

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 13, 2004, 08:39:15 ; Search time 18 Seconds

(without alignments)  
1617.068 Million cell updates/sec

Title: US-09-787-678A-6

Perfect score: 2857  
Sequence: 1 MATRLSEEDNEDNYSLN.....VLDSBALRIRKQSHRIDTDG 559

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	239.5	8.4	369	1	DNAJ_THEMA
2	223	7.8	326	1	DJB6_HUMAN
3	222.5	7.8	376	1	DNAJ2_AQJAE
4	211	7.4	220	1	DJB7_MOUSE
5	209.5	7.3	242	1	DJB6_MOUSE
6	204	7.1	242	1	DJB3_MOUSE
7	196	6.9	280	1	DNAJ_THERM
8	195	6.8	392	1	DNAJ_CHLPP
9	194	6.8	340	1	DJB1_HUMAN
10	193	6.8	348	1	DJB5_MOUSE
11	192.5	6.7	371	1	DNAJ_METIS
12	192.5	6.7	418	1	DNAJ2_ALIPO
13	192	6.7	337	1	DJB4_HUMAN
14	192	6.7	348	1	DJB5_HUMAN
15	191	6.7	334	1	DNAJ1_DROME
16	190	6.7	389	1	DNAJ_MYCSE
17	189	6.6	383	1	DNAJ_LACSK
18	189	6.6	390	1	DNAJ_MYCPN
19	188.5	6.6	307	1	DNAJ_DEIPR
20	188.5	6.6	397	1	DNAJ_MOUSE
21	188	6.6	340	1	DJB1_MOUSE
22	187.5	6.6	373	1	DNAJ_METWA
23	187.5	6.6	397	1	DJB4_HUMAN
24	187	6.5	351	1	DJB2_HUMAN
25	186	6.5	337	1	DJB4_HUMAN
26	185.5	6.5	910	1	DNAJ_MOUSE
27	185	6.5	377	1	DNAJ_PSEAE
28	184.5	6.5	369	1	DNAJ_HELPE
29	184.5	6.5	369	1	DNAJ_NITEU
30	183.5	6.4	419	1	DNAJ_ARATH
31	182.5	6.4	234	1	DNAJ_RHILE
32	182.5	6.4	379	1	DNAJ_LEGPN
33	182	6.4	373	1	DNAJ_CAMOE

34	181	6.3	383	1	DNAJ_PORGI	O9xc66 porphyromon
35	181	6.3	417	1	DNAJ_ATRNU	P43644 atirplex nu
36	180.5	6.3	332	1	DNAJ_STYX3	P50027 syneocyst
37	180.5	6.3	392	1	DNAJ_CHLMO	O9pk53 chlamydia m
38	180.5	6.3	392	1	DNAJ_CHLTR	O84345 chlamydia t
39	180	6.3	377	1	DNAJ_BRUME	O8ye77 bruceella me
40	180	6.3	377	1	DNAJ_BRUSU	O8fx11 bruceella su
41	180	6.3	385	1	DNAJ_CAUCR	P23305 caulobacter
42	180	6.3	387	1	DNAJ_METTS	O9uxr9 methanocarc
43	179.5	6.3	375	1	DNAJ_BROVO	O05980 bruceella ov
44	179.5	6.3	385	1	DNAJ_VIBHA	O87385 vibrio harv
45	179.5	6.3	392	1	NOLC_RHIFR	P26508 rhizobium f

## ALIGNMENTS

RESULT 1	ID	DNAJ_THEMA	STANDARD;	PRT;	369 AA.
AC	O9WZV3;				
DT	16-OCT-2001 (Rel. 40, Created)				
DT	16-OCT-2001 (Rel. 40, Last sequence update)				
DT	15-MAR-2004 (Rel. 43, Last annotation update)				
DE	Chaperone protein dnaJ.				
GN	DNAJ OR TW0849.				
OS	Thermotoga maritima.				
OC	Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.				
OX	NCBI_TaxID=2336;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRATIN=MSB8 / DSM 3109 / ATCC 43589;				
RX	MEDLINE=99287316; Pubmed=10360571;				
RA	Nelson K.E., Clayton R.A., Gill S.R., Gwin M.L., Dodson R.J.,				
RA	Haft D.H., Hickey B.K., Peterson J.D., Nelson W.C., Ketchum K.A.,				
RA	McDonald L., Uetreback T.R., Pratt M.S., Phillips C.A., Richardson D.,				
RA	Stewart A.M., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,				
RA	Heidelberg J., Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.,				
RT	"Evidence for lateral gene transfer between Archaea and Bacteria from				
RL	genome sequence of Thermotoga maritima."				
	Nature 399:323-329(1999).				
CC	-1- FUNCTION: Acts as a co-chaperone. Stimulates, jointly with grpE,				
CC	the ATPase activity of dnaJ (By similarity).				
CC	-1- COFACTOR: Binds 2 zinc ions per monomer (By similarity).				
CC	-1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).				
CC	-1- SIMILARITY: Belongs to the dnaJ family.				
CC	-1- SIMILARITY: Contains 1 CR domain.				
CC	-1- SIMILARITY: Contains 1 CR domain.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>				
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).				
CC	-----				
DR	EMBL; AEO01751; AAD53931.1; --				
DR	PIR; B72327; B72327.				
DR	HSSP; P08622; 1XBL.				
DR	TIGR; TM0849; --				
DR	InterPro; IPR002939; DnaJ_C.				
DR	InterPro; IPR001305; DnaJ_CXXCXGXG.				
DR	InterPro; IPR001623; DnaJ_N.				
DR	InterPro; IPR008971; HSP40_DnaJ_pep.				
DR	InterPro; IPR003095; Hsp_DnaJ.				
DR	Pfam; PF00226; DnaJ_1.				
DR	Pfam; PF01556; DnaJ_C_1.				
DR	Pfam; PF00684; DnaJ_CXXCXGXG; 1.				
DR	PRINTS; PR00625; DNAJPROTEIN.				
DR	SMART; SM00271; DnaJ_1.				
DR	PROSITE; PS00636; DnaJ_1; 1.				



```
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB014888; BAA32209.1; -
DR EMBL; AF075601; AAD3194.1; -
DR EMBL; AF080569; AAD16010.1; ALT_FRAME.
DR EMBL; AB015798; BAA88769.1; -
DR EMBL; AB015799; BAA88770.1; -
DR EMBL; AF060703; AAP21257.1; -
DR EMBL; AL136707; CAB6642.1; -
DR EMBL; BC000177; AAH00177.1; -
DR EMBL; BC002446; AAH02446.1; -
DR HSSP; P25685; 1HDJ.
DR Genew; HGNC:14888; DNABJ6.
DR GO; GO:0003773; P:heat shock protein activity; NAS.
DR InterPro; IPR001623; DnaJ_N.
DR Pfam; PF00226; DnaJ_1.
DR SMART; SM00271; DnaJ_1.
DR PROSITE; PS00636; DnaJ_1; 1.
DR PROSITE; PS50076; DnaJ_2; 1.
DR Chapterone; Alternative splicing.
KM DOMAIN
FT 1 69
FT DOMAIN
FT 83 172
FT VASPLIC 232 241
FT VASPLIC 242 326
FT VASPLIC 326 AA; 36087 MW; ECF1628BF7A524F3 CRC64;
SQ
Query Match 7.8%; Score 223; DB 1; Length 326;
Best Local Similarity 32.1%; Pred. No. 1.1e-08;
Matches 67; Conservative 22; Mismatches 50; Indels 70; Gaps 7;
QY 14 DYSLVLRREASSSEELKAAYRLCMLYHPDKHRDPELKSQAERLFNLVHQAYVLSDP 73
DB 3 DYSLVLRREASSSEELKAAYRLCMLYHPDKHRDPELKSQAERLFNLVHQAYVLSDP 60
QY 74 TRAIYDIYGRKGLMEG-----WEVERRRTPAEIRREFRLQREERRLQORTN 124
DB 61 KRDYIDYGRKGLMEG-----WEVERRRTPAEIRREFRLQREERRLQORTN 103
QY 125 PKGTISGVADTDLFDREDEYEDVSG-----SSPQIEINKMHI 164
DB 104 -----FGGRPFPSFDFPFDFPFDFPFDFPFDFPFDFPFDFPFDFPFDFPFDF 148
QY 165 SOSIAPLTATDTAILG-SLSTONGNGG 192
DB 149 --SFGSGFSPDTGTFTSGSL-----GHGG 171
RESULT 3
DNJ2 AQCAE STANDARD; PRT; 376 AA.
AC 066921;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Chapterone protein dnaJ-2.
GN DNABJ2 OR AQ_703.
OS Aquifex aeolicus.
OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
CX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VF5;
RX MEDLINE=9819666; PubMed=9537320;
Decker G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
```

```
RA Graham D.E., Overbeek R., Sneed M.A., Keller M., Anujay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus."
RL Nature 392:353-358(1998).
CC -1- FUNCTION: Acts as a co-chaperone. Stimulates, jointly with grpe,
CC the ATPase activity of dnaK (By similarity).
CC -1- COFACTOR: Binds 2 zinc ions per monomer (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: Belongs to the dnaJ family.
CC -1- SIMILARITY: Contains 1 J domain.
CC -1- SIMILARITY: Contains 1 CR domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB000703; AAC06881.1; -
DR PIR; E70361; E70361.
DR HSSP; P08622; 1XBL.
DR InterPro; IPR002939; DnaJ_C.
DR InterPro; IPR001305; DnaJ_CXKXGXG.
DR InterPro; IPR001623; DnaJ_N.
DR InterPro; IPR003095; Hsp_DnaJ.
DR Pfam; PF00226; DnaJ_1.
DR Pfam; PF01556; DnaJ_C; 1.
DR Pfam; PF00684; DnaJ_CXKXGXG; 1.
DR PRINTS; PR00625; DnaJPROTEIN.
DR SMART; SM00271; DnaJ_1.
DR PROSITE; PS00636; DnaJ_1; 1.
DR PROSITE; PS50076; DnaJ_2; 1.
DR PROSITE; PS00637; DnaJ_CXKXGXG; FALSE NEG.
KM Chapterone; DNA replication; Heat shock; Repeat; Zinc; Metal-binding;
KM Complete proteome.
FT 8 72
FT DOMAIN
FT REPEAT 156 163
FT REPEAT 172 179
FT REPEAT 194 201
FT REPEAT 207 214
FT METAL 156 156
FT METAL 159 159
FT METAL 172 172
FT METAL 175 175
FT METAL 194 194
FT METAL 197 197
FT METAL 207 207
FT METAL 210 210
SQ SEQUENCE 376 AA; 41963 MW; 9C8F8F829A8A016D CRC64;
Query Match 7.8%; Score 222.5; DB 1; Length 376;
Best Local Similarity 26.4%; Pred. No. 1.5e-08;
Matches 102; Conservative 44; Mismatches 120; Indels 121; Gaps 19;
QY 13 EDYVSLNVRREASSSEELKAAYRLCMLYHPDKHRDPELKSQAERLFNLVHQAYVLSDP 72
DB 7 KDYYEILGVPRNASSQSEIKKAYRLVAKYHPDKCKPE-----CEKFKKEINAEVYVLSDP 62
QY 73 QTRAIYDIYGRKGLMEG-----WEVERRRTP-AEIRREFRLQREERRLQORTN 123
DB 63 EKRLIYDMYGAHAFEGAGAQORVETTEIPRIEILRFPFDFDGSIFERATGRRARRRR 122
QY 124 NPKG-TISVGVDATDLFDREDEYEDVSGSSPQIEINKMHIQSISAPLTATDTAILSG 182
DB 123 SVKGEDEVPEVETL-----LEBAFK--GTVV-IEVER-----EVPSAC----- 159
QY 183 SLSTONGNGGGINFALRLRVTSAKGWELEFGAGDLOGPLFGKLPRLNLPKCFVTNNCA 242
DB 160 -----GGTGYDEBSKSRCTPFTCGGRGE-----TVQGNMF--FQVRQTCFTC----- 197
```

```

QY 243 LQFSSRGIRPGLTTTARNDKNT-VGYLQMRWGIOSAMNTSVPRDTKSHFTVALQGI 301
DB 198 -----GGEVYIENHACTGRGY-----GLVKET-----IKVKI 226
QY 302 PHSFALISYGHKQDDQDTVKSLKGFRC-----TYVEGAERKISR----- 345
DB 227 PPG-----VRDSKLVEGKHAGRGYGGPPGDLTYITVKPKHIFERKDDLYVDV 277
QY 346 -----HSLGAANSVGPQGVSLKVKL 367
DB 278 NITYPEAVLCTEVEVPTLDEKXKVKI 304

RESULT 4
DJB7 MOUSE STANDARD; PRT; 220 AA.
ID DJB7 MOUSE
AC 09OY18;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DnaJ homolog subfamily B member 7 (mdj5).
GN DnaJb7.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=CD-1;
RX MEDLINE=21023480; PubMed=11147971;
RT Ohtsuka K., Hata M.;
RT "Mammalian HSP40/DNAJ homologs: cloning of novel cDNAs and a proposal
RT for their classification and nomenclature.";
RL Cell Stress Chaperones 5:98-112(2000).
CC -1- SIMILARITY: Contains 1 J domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB028855; BAA8303.1; -.
DR HSPB; P25685; 1HDJ.
DR MGD; MGI:1914012; DnaJb7.
DR InterPro; IPR001623; DnaJ_N.
DR Pfam; PF00226; DnaJ; 1.
DR PROSITE; PS00636; DnaJ_1; 1.
DR PROSITE; PS50076; DnaJ_2; 1.
DR Chapterone.
FT DOMAIN 3 69 J-DOMAIN.
SQ SEQUENCE 220 AA; 25530 MW; E741D9232713F7E CRC64;

Query Match 7.4%; Score 211; DB 1; Length 220;
Best Local Similarity 33.3%; Pred. No. 4.7e-08;
Matches 59; Conservative 25; Mismatches 49; Indels 44; Gaps 6;

QY 14 DYISLVNRRASSEELKAAVRLCMTHPKRDPELKSQARLRLVHQAVEYSDPQ 73
DB 3 DYIEVGLGVORYAPEDIKRAVRYVALKMHDPK--NPNNEKEARKEKVAEAAYEVSINVE 60
QY 74 TRAIYDIYGRKGLMEGMEVEVRRRPAEIRFEERLQREERERRLQQRTPNGTISVGV 133
DB 61 KRRIYDKYKGGGLDGRG-----ASHLDREKER----- 88
QY 134 DATDLEDRYDEEYEDVSGSSFPQIEINKMI-SQSIAPPLATATDAILSGSLSTONG 189
DB 89 ---FTFRKADVDFEKEIFGERD---FSFLLFEDSLGGLINSRTP--SGSRERRRAG 136
RESULT 5

```

```

DJB6 MOUSE STANDARD; PRT; 242 AA.
ID DJB6 MOUSE
AC 054946; 09OY19;
DT 30-MAY-2000 (Rel. 39, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE DnaJ homolog subfamily B member 6 (Heat shock protein J2) (HSJ-2)
DE (MRJ) (mdj4).
GN DnaJb6 OR HSJ2 OR MRJ.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=129; TISSUE=Trophoblast;
RL Hunter P.J., Swanson B.U., Haendel M., Lyons G.E., Crose J.C.;
RL Submitted (May-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6J;
RX MEDLINE=21023480; PubMed=11147971;
RT Ohtsuka K., Hata M.;
RT "Mammalian HSP40/DNAJ homologs: cloning of novel cDNAs and a proposal
RT for their classification and nomenclature.";
RL Cell Stress Chaperones 5:98-112(2000).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=22388257; PubMed=12477932;
RX Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RX Klausner R.D., Collins F.S., Wagner L., Scheffen C.M., Schuler G.D.,
RX Altschuler S.F., Zeeberg B., Bueto K.H., Schaefer C.F., Bhat N.K.,
RX Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RX Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RX Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RX Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RX Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullihy S.J.,
RX Bosak S.A., Mcwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RX Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Huiyk S.W.,
RX Villalon D.K., Murry D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RX Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RX Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RX Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RX Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RX Butlerfield V.S.N., Krzywinski M.I., Skalska U., Snailus D.E.,
RX Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -1- SIMILARITY: Contains 1 J domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF035962; AAC16759.1; -.
DR EMBL; AB028854; BAA8302.1; -.
DR EMBL; BC003702; AAH03702.1; -.
DR HSPB; P25685; 1HDJ.
DR MGD; MGI:1344381; DnaJb6.
DR InterPro; IPR001623; DnaJ_N.
DR Pfam; PF00226; DnaJ; 1.
DR SMART; SM00271; DnaJ; 1.
DR PROSITE; PS00636; DnaJ_1; 1.
DR PROSITE; PS50076; DnaJ_2; 1.
DR Chapterone.
FT DOMAIN 1 69 J-DOMAIN.
FT DOMAIN 83 172 GUY/PHE-RICH.
FT CONFLICT 139 140 GS -> AP (IN REF. 1).

```



```

FT CONFLICT 218 218 E -> V (IN REF. 3).
FT CONFLICT 227 227 S -> P (IN REF. 1).
SQ SEQUENCE 242 AA; 26978 MM; 3856053CJBB0735D CRC64;

Query Match
Best Local Similarity 31.0%; Score 209.5; DB 1; Length 242;
Pred. No. 6.8e-08;
Matches 65; Conservative 21; Mismatches 53; Indels 71; Gaps 7;

QY 14 DYSLNVRREASSEELKAVRRLCMLYHPDKRDPDLKSAERLFLVHQAYEVLSDPQ 73
DB 3 DYVVLGVQORASPEDIKKAYRKQALKWHDPK--NPEKKEAEKRFQVAVAEVLSDAK 60
QY 74 TRAIYDIYGRKGLMEG-----WEVERRRTPAIRERERLQREERERLQORT 123
DB 61 KRDIYDKYKKGKGLNGGGGGGHHDSPEFGFTFRNPDDVREF----- 104
QY 124 NPKGTISGVDTATDLFPRYDEYEDVSG-----SFPQIEINM 163
DB 105 -----FGGRDPSFDFDFDFDFGNNRGRGNSRGAGSPFTSPSGFP----- 149
QY 164 ISQIEAPLTATDTAALS-GSLSTONGCG 192
DB 150 ---SFGSGFPAFDGTFTPGSL---GKG 172

RESULT 6
DJB3_MOUSE STANDARD; PRT; 242 AA.
ID DJB3_MOUSE
AC 035723; Q9DAN3; Q9DAN4;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE DnaJ homolog subfamily B member 3 (DnaJ protein homolog 3) (Heat shock
U3 protein) (HSC-3) (MSJ-1).
GN DNABJ3 OR HSC3 OR MSJ1.
OC Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=98189155; PubMed=9521861;
RA Berruti G., Perigo L., Borgonovo B., Martegani E.;
RT "MSJ-1, a new member of the DnaJ family of proteins, is a male germ
cell-specific gene product.";
RL Exp. Cell Res. 239:430-441(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Aizawa K., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
RA Saito T., Okazaki Y., Gojohori T., Bono H., Kanukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner T., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochwa H.,
RA Knehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schirral L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okada T., Furuno M., Aono H., Baldarelli I., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gaitanaris M.,
RA Guernicich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaeerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Saeki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilmink L.,
RA Wyszewski-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohlschki S.,
RA Hayashizaki Y.;
RA "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
CC -1- FUNCTION: May operate as a co-chaperone of the male germ cell- and
haploid stage-specific Hsp70 proteins.

```

```

CC -1- TISSUE SPECIFICITY: Testis specific. Expression is confined to the
CC germ line without any contribution of the somatic components.
CC -1- DEVELOPMENTAL STAGE: Its expression occurs in the postembryonic
CC phase of male germ cell development. First detected in 30 days old
CC mice and thereafter into adulthood. Barely detectable in 20 days
CC old mice and absent before this period.
CC -1- SIMILARITY: Contains 1 J domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U95607; AAC13944.1; -.
DR EMBL; AK005690; BAB24188.1; -.
DR EMBL; AK005688; BAB24186.1; -.
DR HSSP; P25685; 1HDJ.
DR MGD; MGI:1306822; DnaJb3.
DR InterPro; IPR001623; DnaJ_N.
DR Pfam; PF00226; DnaJ_1.
DR SMART; SM00271; DnaJ_1.
DR PROSITE; PS00636; DnaJ_1; 1.
DR PROSITE; PS50076; DnaJ_2; 1.
KM Chapterone.
FT DOMAIN 1 69 J-DOMAIN.
FT CONFLICT 71 71 V -> M (IN REF. 2; BAB24186).
FT CONFLICT 149 149 F -> I (IN REF. 2; BAB24188).
FT CONFLICT 233 233 R -> K (IN REF. 2).
SQ SEQUENCE 242 AA; 26678 MM; 5F54P48A00094D84 CRC64;

Query Match
Best Local Similarity 30.6%; Score 204; DB 1; Length 242;
Pred. No. 1.7e-07;
Matches 63; Conservative 31; Mismatches 64; Indels 48; Gaps 7;

QY 14 DYSLNVRREASSEELKAVRRLCMLYHPDKRDPDLKSAERLFLVHQAYEVLSDPQ 73
DB 3 DYVVLGVQORASPEDIKKAYRKQALKWHDPK--NPEKKEAEKRFQVAVAEVLSDAK 60
QY 74 TRAIYDIYGRKGLMEG-----WEVERRRTPAIRERERLQREERERLQORT 108
DB 61 KREYIDKCGVGV-EVGGGGAAGSPFHDAPFYVSFPDPAVRFEPFGHDPSPDFPGCD 119
QY 109 RLOREERERLQORTPKGTISGVDTATDLFPRYDEYEDVSGSSPQIEINKMHSQST 168
DB 120 PLENFEGDRSTRSGSRNRGAVPPTSPTE-FPGF-----GGGF-----ASL 159
QY 169 EAPLTATDTAALSGSLSTONGCGS 194
DB 160 DTGFTSFGSPNGSLSPMSGCGGA 185

RESULT 7
DnaJ_THETH STANDARD; PRT; 280 AA.
ID DnaJ_THETH
AC Q56237; P77642;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Chapterone protein dnaJ.
GN DnaJ.
OS Thermus thermophilus.
OC Bacteria; Deinococcus-Thermus; Deinococci; Thermales; Thermaceae;
OC Thermus.
OX NCBI_TaxID=274;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HB8 / ATCC 27634;
RX MEDLINE=98007877; PubMed=9349721;
RA Osipjuk J., Joachimiak A.;
RA "Cloning, sequencing, and expression of dnaJ-operon proteins from the

```

RT thermophilic bacterium *Thermus thermophilus*.";  
RL Biochim. Biophys. Acta 1353:253-265(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RR STRAIN=HB8 / ATCC 27634;  
RA Seidel R.;  
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RR STRAIN=HB8 / ATCC 27634;  
RX MEDLINE=97420496; PubMed=9276481;  
RA Mochizashi K., Yoda M., Oda K., Yoshida M.;  
RL "K+ is an indispensable cofactor for GTPe stimulation of Arpase  
RT activity of DnaK/DnaJ complex from *Thermus thermophilus*.";  
RS FEBS Lett. 412:633-636(1997).  
RN [4]  
RP FUNCTION.  
RX MEDLINE=99194855; PubMed=10092456;  
RA Klostermeier D., Seidel R., Reinsteil J.;  
RL "The functional cycle and regulation of the *Thermus thermophilus* DnaK  
RT chaperone system.";  
RS J. Mol. Biol. 287:511-525(1999).  
RL -I- FUNCTION: Does not influence ATP binding or hydrolysis nor ADP  
CC release. Exerts influence on the interaction of DnaK with  
CC substrates; in the presence of Dafa, DnaJ inhibits substrate  
CC binding, and substrate already bound to DnaK is displaced by DnaJ  
CC and Dafa.  
CC  
CC -I- SUBUNIT: Forms a heteronamer with DnaJ and Dafa in the resting  
CC state. Three copies of each protein are present in the complex.  
CC -I- SUBCELLULAR LOCATION: Cytoplasmic (by similarity).  
CC -I- SIMILARITY: Belongs to the dnaJ family.  
CC  
CC -I- SIMILARITY: Contains 1 J domain.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation --  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL, L57504; AAA04678.1; -.  
DR EMBL, Y07826; CA69161.1; -.  
DR EMBL, D84222; BA1282.1; -.  
DR EMBL, AB012390; BAA81743.1; -.  
DR EMBL, AB032368; BAA96087.1; -.  
DR HSSP; P25685; 1HDJ.  
DR InterPro; IPR002939; DnaJ\_C.  
DR InterPro; IPR001623; DnaJ\_N.  
DR InterPro; IPR003095; Hep\_DnaJ.  
DR Pfam; PF00226; DnaJ\_1.  
DR Pfam; PF01556; DnaJ\_C\_1.  
DR PRINTS; PR00625; DNAJPROTEIN.  
DR SMART; SMO0271; DnaJ\_1.  
DR PROSITE; PS00636; DNAJ\_1; 1.  
DR PROSITE; PS50076; DNAJ\_2; 1.  
KW Chaperone; DNA replication; Heat shock.  
FT DOMAIN 4  
FT DOMAIN 72  
FT DOMAIN 75  
FT DOMAIN 80  
FT POLY-PRO.  
FT CONFLICT 129 129 E -> K (IN REF. 1).  
SQ SEQUENCE 280 AA; 30978 MW; AE32D59535388BD CRC64;

```

Query Match      6 9%  Score 196;  DB 1;  Length 280;
Best Local Similarity 36.2%  Pred. No. 7.7e-07;
Matches 46;  Conservative 20;  Mismatches 31;  Indels 30;  Gaps 4;

QY      13 EDVYSLNVRREASEEELKAAVRRRLCMVHPDKRHPDLKSQAERLFNLVHQAAVEVLSDP 72
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DQ      5 KDYVAILGVPRNNAQGEETKRAYKRLRLARGYHDVNVKSP-----AESEKFEINEAVALISDP 60
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

QY      73 QTRALVYDYGK-----GLEMEGEEVVERKRRTPMEIEEERRL-----QRE 113
      :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

Db 61 EKRRIDVYGTTEAPPPPPPGSYDSRSDV-----EDISEFFQELFGFGLGFGFRRS 113

Qy 114 REERRIQ 120  
|:|:|:  
114 RKGRDLR 120

ID	NAME	STANDARD	PR	AA
DT	16-OCT-2001	Rel. 40, Created)		
DT	16-OCT-2001	Rel. 40, Last sequence update)		
DT	10-OCT-2003	Rel. 42, Last annotation update)		
DE	Chaperone protein dnaJ			
GN	DNAJ OR CPN0032 OR CP0744 OR CPB0036			
OS	Chlamydia pneumoniae (Chlamydia pneumoniae)			
OC	Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydiales			
OX	NCBI_TaxID=83558;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CWL029;			
RC	MEDLINE=99206606; PubMed=10192388;			
RA	Kalman S., Mitchell W., Marache R., Lammell C., Fan J., Hyman R.W.,			
RA	Olinger U., Grimwood J., Davis R.W., Stephens R.S.; trachomatis."			
RT	Comparative genomes of Chlamydia pneumoniae and C. trachomatis."			
RL	Nat. Genet. 21:385-389(1999).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=AR39;			
RC	MEDLINE=20150255; PubMed=10684935;			
RA	Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F., Bass S.,			
RA	White O., Hickey E.K., Peterson J., Ulfersack T., Berry K., Bass S.,			
RA	Linher K., Weidman J., Knouri H., Craven B., Bowman C., Dodson R.,			
RA	Gwin M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,			
RA	Eisen J., Fraser C.M.;			
RT	"Genome sequences of Chlamydia trachomatis Moyn and Chlamydia			
RT	pneumoniae AR39."			
RL	Nucleic Acids Res. 28:1397-1406(2000).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=J138;			
RC	MEDLINE=20330349; PubMed=10871362;			
RA	Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,			
RA	Shiba T., Ishii K., Hattori M., Kunara S., Nakazawa T.;			
RT	"Comparison of whole genome sequences of Chlamydia pneumoniae J138			
RT	from Japan and CWL029 from USA."			
RL	Nucleic Acids Res. 28:2311-2314(2000).			
RN	[4]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=TW-183;			
RA	Geng M.M., Schumacher A., Muehldorfer I., Bensch K.W., Schaefer K.P.,			
RA	Schneider S., Pohl T., Baeig A., Murre R., Melchers K.;			
RT	"The genome sequence of Chlamydia pneumoniae TW183 and comparison with			
RT	other Chlamydia strains based on whole genome sequence analysis."			
RL	Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.			
CC	-1- FUNCTION: Acts as a co-chaperone. Stimulates, jointly with grpE,			
CC	the Arpase activity of dnaK (By similarity).			
CC	-1- COFACTOR: Binds 2 zinc ions per monomer (By similarity).			
CC	-1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).			
CC	-1- SIMILARITY: Belongs to the dnaJ family.			
CC	-1- SIMILARITY: Contains 1 CR domain.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
OR	EMBL, AB001588; AAD18185.1; --			

```

DR EMBL; AE002233; AAF38549.1; -
DR EMBL; AP002545; BAA98244.1; -
DR EMBL; AE017157; AAP97969.1; -
DR PIR; B66495; B66495.
DR PIR; G71218; G71218.
DR HSSP; P08622; 1XBL.
DR PHCI-2DPAGE; Q929E9; -.
DR TIGR; CP0744; -.
DR InterPro; IPR002939; DnaJ_C.
DR InterPro; IPR001305; DnaJ_CXKXGK.
DR InterPro; IPR001623; DnaJ_N.
DR InterPro; IPR008971; HSP40_DnaJ_pdp.
DR InterPro; IPR003095; Hsp_DnaJ.
DR Pfam; PF00226; DnaJ_1.
DR Pfam; PF01556; DnaJ_C_1.
DR Pfam; PF00684; DnaJ_CXKXGK; 1.
DR PRINTS; PRO0625; DnaJPROTEIN.
DR SMART; SM00271; DnaJ_1.
DR PROSITE; PS00636; DnaJ_1; 1.
DR PROSITE; PS50076; DnaJ_2; 1.
DR PROSITE; PS00637; DnaJ_CXKXGK; FALSE_NEG.
KM Chaperone; DNA replication; Heat shock; Repeat; Zinc; Metal-binding;
KM Complete proteome.
FT DOMAIN 2 67 J-DOMAIN.
FT REPEAT 161 128 GLY-RICH.
FT REPEAT 178 168 CXKXGK MOTIF.
FT REPEAT 200 207 CXKXGK MOTIF.
FT REPEAT 214 221 CXKXGK MOTIF.
FT METAL 161 161 ZINC 1 (BY SIMILARITY).
FT METAL 164 164 ZINC 1 (BY SIMILARITY).
FT METAL 178 178 ZINC 2 (BY SIMILARITY).
FT METAL 181 181 ZINC 2 (BY SIMILARITY).
FT METAL 200 200 ZINC 2 (BY SIMILARITY).
FT METAL 203 203 ZINC 2 (BY SIMILARITY).
FT METAL 214 214 ZINC 1 (BY SIMILARITY).
FT METAL 217 217 ZINC 1 (BY SIMILARITY).
SQ SEQUENCE 392 AA; 42108 MW; 28BDAB99C4E1A531 CRC64;

Query Match 6.8%; Score 195; DB 1; Length 392;
Best Local Similarity 25.8%; Pred. No. 1.4e-06;
Matches 100; Conservative 51; Mismatches 140; Indels 96; Gaps 21;

14 DYSLANVRASSEELKAAVRLCMYHPDKHRDPELKSQAEELFNVMQAVVLSDPQ 73
2 DYSLIGISKTASAEIKKATRKLAIVKHPDKNPG--DAAAEKRFKVEVSAYEVLSDPQ 58
74 TRAIYDIYKKRGL-----EMEGMEVVERRRTPAIREFEERLQREERRRL 119
59 KRDSYDRFGKDGFPAGAGGFGGAGMGNGME-----DALRT--FMGAF----- 98
120 QQRNPNGKTSIVGDADLDLPRYREY--EDVGSSEFPQIEINKMHISQIEAPLATYD 176
99 -----GGEFGGSGFFDGLFGCLGEAFGRSDPAGAR--QASKKVHINLTFEEAAGVE 150
177 -TALLSGSLSTQNONGGSSIN-FALRRYTSKKGCELEFGAGDLOGLPLGKLFRNLPR 234
151 KEIVVSGKSCETSGGQAVVPQIGKSCERCKSGGQVQVOSRG-----FFSN--ASTCPE 202
235 C-----FVTTNCALOFSRGIRPGLTTLVARNLDKNTVQLQPMRGISQAMNTSIVBDTK 289
203 CCGSGRIITDPCS---SGRG-----QGRVNDKRSV-NHVIHAGVDSGR-----LK 244
290 TSHETVALQIGIP---HSPALISYQHKQ--DDQTRVKSLKAGFGTVVEYGARKI 343
245 MEGYGDAGQNGAPBGDLYVFIDVESHVFERRRDILL--ELPIGFVDAL--GMKKEI 299
344 SRHSVLAASVGVPGVGS-----LKVX 366
Db 300 PTLTKGSGRLTVBGIQSGTTLIKVR 326

```

```

DJB1_HUMAN
ID DJB1_HUMAN STANDARD; PRT; 340 AA.
AC P25685;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE DnaJ homolog subfamily B member 1 (Heat shock 40 kDa protein 1) (Heat
DE shock protein 40) (HSP40) (DnaJ protein homolog 1) (HDJ-1).
GN DNABJ1 OR HSPF1 OR DNaJ1 OR HDJ1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_Taxid=9606;
[1]
RN 1;
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=92091635; PubMed=1754405;
RA Raabe T., Manley J.L.;
RT "A human homologue of the Escherichia coli DnaJ heat-shock protein.";
RL Nucleic Acids Res. 19:6645-6645(1991).
[2]
RN 2;
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-48.
RC TISSUE=Placenta;
RX MEDLINE=94071949; PubMed=8250930;
RA Ohtsuka K.;
RT "Cloning of a cDNA for heat-shock protein hsp40, a human homologue of
RT bacterial DnaJ.";
RL Biochem. Biophys. Res. Commun. 197:235-240(1993).
[3]
RN 3;
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=97131529; PubMed=8975727;
RA Hata M., Okumura K., Seto W., Ohtsuka K.;
RT "Genomic cloning of a human heat shock protein 40 (Hsp40) gene (HSPF1)
RT and its chromosomal localization to 19p13.2.";
RL Genomics 38:446-449(1996).
[4]
RN 4;
RP SEQUENCE FROM N.A.
RC TISSUE=Brain, and Lung;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.T., Wang J., Heien F.,
RA Dackiwenko L., Marusha K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Buttefield Y.S.N., Krzywinski M.T., Skalska U., Smalins D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
[5]
RN 5;
RP STRUCTURE BY NMR OF 1-76.
RX MEDLINE=96291433; PubMed=8764402;
RA Qian Y.Q., Patel D., Harcl F.-U., McCoil D.J.;
RT "Nuclear magnetic resonance solution structure of the human Hsp40
RT (HDJ-1) J-domain";
RL J. Mol. Biol. 260:224-235(1996).
CC -!- FUNCTION: Interacts with HSP70 and can stimulate its ATPase
CC activity. Stimulates the association between HSC70 and Hsp.
CC -!- SIMILARITY: Contains 1 J domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; X62421; CA444287.1; -  
 DR EMBL; D4547; BAA08495.1; -  
 DR EMBL; D8429; BAA12819.1; -  
 DR EMBL; BC002352; AAH02352.1; -  
 DR EMBL; BC019827; AAH19827.1; -  
 DR PIR; JN0912; JN0912.  
 DR PIR; S20062; S20062.  
 DR PDB; 1HDJ; 08-NOV-96.  
 DR Genew; HGNC:5270; DNABJ1.  
 DR MIM; 604572; -  
 DR GO; 0003773; F:heat shock protein activity; TMS.  
 DR InterPro; IPR002939; DnaJ\_C.  
 DR InterPro; IPR001623; DnaJ\_N.  
 DR InterPro; IPR008971; Hsp40 DnaJ\_pep.  
 DR InterPro; IPR003095; Hsp\_DnaJ.  
 DR Pfam; PF00226; DnaJ\_1.  
 DR Pfam; PF01556; DnaJ\_C\_1.  
 DR PRINTS; PR00625; DnaJPROTEIN.  
 DR SMART; SM00271; DnaJ\_1.  
 DR PROSITE; PS00636; DnaJ\_1; 1.  
 DR PROSITE; PSS0076; DnaJ\_2; 1.  
 DR Heat shock; Chaperone; 3D-structure.  
 KM DOMAIN 1 70  
 FT DOMAIN 1 70  
 FT CONFLICT 11 11  
 FT CONFLICT 13 28  
 FT CONFLICT 13 28  
 FT CONFLICT 68 68  
 FT CONFLICT 81 136  
 FT CONFLICT 150 150  
 FT CONFLICT 183 183  
 FT CONFLICT 320 320  
 FT HELIX 6 9  
 FT TURN 10 10  
 FT TURN 13 14  
 FT HELIX 17 29  
 FT TURN 30 31  
 FT TURN 33 35  
 FT TURN 39 40  
 FT HELIX 41 54  
 FT TURN 55 56  
 FT HELIX 58 66  
 FT TURN 67 67  
 FT HELIX 69 71  
 SQ SEQUENCE 340 AA; 38044 MW; 17545098B0C196DF CRC64;

Query Match 6.8%; Score 194; DB 1; Length 340;  
 Best Local Similarity 48.7%; Pred. No. 1.4e-06;

Matches 38; Conservative 14; Mismatches 22; Indels 4; Gaps 1;

QY 13 EDVYSLNVRRAASEETKAAVRLCLYPPDKGRPELKSQAEKRLFNLVHQAIVTLSP 72  
 DB 3 KDYYQTGLARGASDEIRKAVRRQALRYHPDKNEP---GAEEKFEIAEAAYDVLSIP 58  
 QY 73 QTRAIYDIYGRGLIEMEG 90  
 DB 59 RKREIFPRYGEGLKSG 76

RESULT 10  
 DUBS\_MOUSE STANDARD; PRT; 348 AA.  
 AC 089174;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE DnaJ protein subfamily B member 5 (Heat shock protein Hsp40-3) (Heat  
 DE shock protein cognate 40) (Hsc40).  
 GN DNABJ5 OR HSC40.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_Taxid=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20035738; PubMed=10570961;  
 RT Chen M.-S., Roti J.R., Laszlo A.;  
 RT "Hsc40, a new member of the hsp40 family, exhibits similar expression  
 RT profile to that of hsc70 in mammalian cells.";  
 RL Gene 238:333-341(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6; TISSUE=Brain;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RX Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Wax S.I., Wang J., Hsieh F.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loughran N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
 RA Bosak S.A., McKernan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting R.W., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butlerfield Y.S.N., Krzywinski M.I., Skalska U., Smalins D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 CC -1- INDUCTION: Expressed under normal conditions, its expression can  
 CC further be increased after various stress treatments.  
 CC -1- SIMILARITY: Contains 1 J domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; AF092535; AAC64141.1; -  
 DR EMBL; AF088983; AAC35861.1; -  
 DR EMBL; AF321322; AAG53972.1; -  
 DR EMBL; BC057087; AAH57087.1; -  
 DR HSSP; P25685; 1HDJ.  
 DR MGD; MGI:1930018; DnaJb5.  
 DR InterPro; IPR002939; DnaJb5.  
 DR InterPro; IPR001623; DnaJ\_C.  
 DR InterPro; IPR008971; Hsp40\_DnaJ\_pep.  
 DR InterPro; IPR003095; Hsp\_DnaJ.  
 DR Pfam; PF00226; DnaJ\_1.  
 DR Pfam; PF01556; DnaJ\_C\_1.  
 DR PRINTS; PR00625; DnaJPROTEIN.  
 DR PROSITE; PS00636; DnaJ\_1; 1.  
 DR PROSITE; PSS0076; DnaJ\_2; 1.  
 KM Chaperone.  
 FT DOMAIN 4 68  
 FT SEQUENCE 348 AA; 39119 MW; 833D2F3547321687 CRC64;

Query Match 6.8%; Score 193; DB 1; Length 348;  
 Best Local Similarity 43.6%; Pred. No. 1.7e-06;  
 Matches 34; Conservative 20; Mismatches 20; Indels 4; Gaps 1;

QY 13 EDVYSLNVRREASSEELKAAYRRLCM,YHPDKRDELPKLSQAEKRLFNLVHQAAYEVLSDP 72  
DB 3 KQYKXICLIPSGANDEDEIKAIKRIKALKTHPPKNEP---NAEKFEIAMEAYDVLSDP 58  
QY 73 QTRAIDYDYGKRGLEMEG 90  
DB 59 KKRSLYDQYGGELKTG 76

RESULT 11  
ID DNJ2\_METSS STANDARD; PRT; 371 AA.  
AC Q92FC5;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Chaperone protein dnaJ.  
GN DNJ2.  
OS Methylovorus sp. (strain S81 / DSM 11726).  
OC Bacteria; Proteobacteria; Betaproteobacteria; Methylophilales;  
OC Methylophilaceae; Methylovorus.  
OX NCBI\_TaxID=81683;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Eom C.Y., Kim Y.M.;  
RT "grpe, dnaK, and dnaJ genes of Methylovorus sp. strain S81 DSM11726";  
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
CC -1 FUNCTION: Acts as a co-chaperone. Stimulates, jointly with grpe,  
the ATPase activity of dnaK (By similarity).  
CC -1 COPACITOR: Binds 2 zinc ions per monomer (By similarity).  
CC -1 SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
CC -1 SIMILARITY: Belongs to the dnaJ family.  
CC -1 SIMILARITY: Contains 1 CR domain.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AF106835; AAC95379.1; --  
DR HSSP; P08622; 1BQZ.  
DR InterPro; IPR002939; DnaJ\_C.  
DR InterPro; IPR001305; DnaJ\_CXXCXGKG.  
DR InterPro; IPR001623; DnaJ\_N.  
DR InterPro; IPR003095; Hsp\_DnaJ.  
DR Pfam; PF00226; DnaJ\_1.  
DR Pfam; PF01556; DnaJ\_C; 1.  
DR Pfam; PF00684; DnaJ\_CXXCXGKG; 1.  
DR PRINTS; PR00625; DNJAPROTEIN.  
DR SMART; SMO0271; DnaJ\_1.  
DR PROSITE; PS00636; DnaJ\_1; 1.  
DR PROSITE; PS0076; DnaJ\_2; 1.  
DR Chaperone; DNA replication; Heat shock; Repeat; Zinc; Metal-binding.  
KW Chaperone; DNA replication; Heat shock; Repeat; Zinc; Metal-binding.  
FT DOMAIN 5 70  
FT DOMAIN 70  
FT REPEAT 145 152 CXXCXGKG MOTIF.  
FT REPEAT 162 169 CXXCXGKG MOTIF.  
FT REPEAT 184 191 CXXCXGKG MOTIF.  
FT REPEAT 198 205 CXXCXGKG MOTIF.  
FT METAL 145 145 ZINC 1 (BY SIMILARITY).  
FT METAL 148 148 ZINC 1 (BY SIMILARITY).  
FT METAL 162 162 ZINC 2 (BY SIMILARITY).  
FT METAL 165 165 ZINC 2 (BY SIMILARITY).  
FT METAL 184 184 ZINC 2 (BY SIMILARITY).  
FT METAL 187 187 ZINC 1 (BY SIMILARITY).  
FT METAL 198 198 ZINC 1 (BY SIMILARITY).  
FT METAL 201 201 ZINC 1 (BY SIMILARITY).

SQ SEQUENCE 371 AA; 41052 MW; 86C1AF5EF87CF95C CRC64;  
Query Match 6.7%; Score 192.5; DB 1; Length 371;  
Best Local Similarity 50.7%; Pred. No. 2e-06;  
Matches 38; Conservative 12; Mismatches 22; Indels 3; Gaps 1;  
QY 13 EDVYSLNVRREASSEELKAAYRRLCM,YHPDKRDELPKLSQAEKRLFNLVHQAAYEVLSDP 72  
DB 4 KQYKXICLIPSGANDEDEIKAIKRIKALKTHPPKNEP---NPKAESEIKAEKAEVLSDE 60  
QY 73 QTRAIDYDYGKRGLE 87  
DB 61 QKRAYDYGGHAGVD 75

RESULT 12  
ID DNJ2\_ALLPO STANDARD; PRT; 418 AA.  
AC P42824;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE DnaJ protein homolog 2.  
GN LDJ2.  
OS Allium porrum (leek).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Asparagales; Alliaceae;  
OC Allium.  
OX NCBI\_TaxID=4681;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Bessoule J.J., Teset E., Casagane C.;  
RT "Cloning of a new isoform of a DnaJ protein from Allium porrum  
epidermal cells.";  
RL Plant Physiol. Biochem. 32:723-727(1994).  
CC -1 FUNCTION: Have a continuous role in plant development probably in  
the structural organization of compartments (By similarity).  
CC -1 SUBCELLULAR LOCATION: Membrane-bound (potential).  
CC -1 SIMILARITY: Contains 1 CR domain.  
CC -1 SIMILARITY: Contains 1 CR domain.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X77632; CA54720.1; --  
DR PIR; S42031; S42031.  
DR HSSP; P25685; 1HDV.  
DR InterPro; IPR002939; DnaJ\_C.  
DR InterPro; IPR001305; DnaJ\_CXXCXGKG.  
DR InterPro; IPR001623; DnaJ\_N.  
DR InterPro; IPR008971; HSP40\_DnaJ\_pcp.  
DR InterPro; IPR003095; Hsp\_DnaJ.  
DR Pfam; PF00226; DnaJ\_1.  
DR Pfam; PF01556; DnaJ\_C; 1.  
DR Pfam; PF00684; DnaJ\_CXXCXGKG; 1.  
DR PRINTS; PR00625; DNJAPROTEIN.  
DR SMART; SMO0271; DnaJ\_1.  
DR PROSITE; PS00636; DnaJ\_1; 1.  
DR PROSITE; PS00637; DnaJ\_CXXCXGKG; 1.  
DR PROSITE; PS0076; DnaJ\_2; 1.  
DR Chaperone; Repeat; Prenylation; Lipoprotein; Multigene family.  
FT DOMAIN 11 76  
FT DOMAIN 76  
FT DOMAIN 83 86 POLY-GLY.  
FT DOMAIN 81 108 GLY-RICH.  
FT REPEAT 148 155 CXXCXGKG MOTIF.  
FT REPEAT 164 171 CXXCXGKG MOTIF.  
FT REPEAT 191 198 CXXCXGKG MOTIF.  
FT REPEAT 207 214 CXXCXGKG MOTIF.

FT LIPID 415 415 S-farnesyl cysteine (By similarity).  
 SQ SEQUENCE 418 AA; 46584 MW; DCE2A4DF19232956 CRC64;  
 Query Match 6.7%; Score 192.5; DB 1; Length 418;  
 Best Local Similarity 29.4%; Pred. No. 2,4e-06;  
 Matches 65; Conservative 30; Mismatches 69; Indels 57; Gaps 9;  
 QY 11 DNEEDYLLNVRREASSEELKAAYRRLCMLYHPDKHDPDKSQARLFWLHQAAYVLS 70  
 10 DNTKYIVLVGSKATPEDLKAYRKALIKHPDKGDPE-----KFEIGQAEVLN 62  
 QY 71 DPQTRAIYDIYKRGKLEMEGMEVEVRRRTPAIREEF-----ERLQREERERLQQR 122  
 63 DPEKREIYDYGEGRLK-EGMGCGGVHDPFDIRQSPFGGSGGSSRGRRQRGEDEV 121  
 QY 123 TNKRGITSVGVADATDLPDRIDEDEVYSGSSFOIEINKKHISQISAPLTADTALISG 182  
 122 VHP-----LKVSELDLY-----NGTS-----KKLISLRN-----VLCT 149  
 QY 183 SLSTONGGSGSINFALRVTSAGKMGCE-----LEFGAGDLO 219  
 Db 150 KCKGKSGSKSGAMNCA-----SCGSGMKYSIRQLGFGMIQ 185  
 RESULT 13  
 DJB4 HUMAN STANDARD; PRT; 337 AA.  
 AC Q9UDF4; Q13431; 40, Created  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE DnaJ homolog subfamily B member 4 (Heat shock 40 kDa protein 1 homolog) (Heat shock protein 40 homolog) (Hsp40 homolog).  
 GN DNABJ4 OR HJL1 OR DNABJ4.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OC NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Liver;  
 RX MEDLINE=9820766; PubMed=9546042;  
 RA Krausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Sherman C.M., Schuler G.D., Altschul S.F., Zeeberg B., Bietow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heileh F., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C., Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J., Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W., Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butcherfield Y.S.N., Krzyzanski M.I., Skalska U., Smalins D.E., Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.; "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences".  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RP [3]  
 PRELIMINARY SEQUENCE FROM N.A.

RC TISSUE=Liver;  
 RA Won M., Moon K.M., Lee C.E., Yoo H.S.;  
 RL Submitted (NOV-1995) to the EMBL/GenBank/DDA databases.  
 CC -1- INDUCTION: By heat shock.  
 CC -1- SIMILARITY: Contains 1 J domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; UA0992; AAC14483.2; -  
 DR EMBL; BC034721; AAH34721.1; -  
 DR EMBL; U41290; AAB07346.1; ALT\_FRAME.  
 DR HSSP; P25685; 1HDJ.  
 DR Genew; HGNC:14886; DNABJ4.  
 DR GO; GO:0003773; P:heat shock protein activity; TAS.  
 DR GO; GO:0009408; P:response to heat; TAS.  
 DR InterPro; IPR002939; DnaJ\_N.  
 DR InterPro; IPR001623; DnaJ\_C.  
 DR InterPro; IPR008971; Hsp40\_DnaJ\_pdp.  
 DR Pfam; PF00226; DnaJ\_1.  
 DR Pfam; PF01566; DnaJ\_C\_1.  
 DR PRINTS; PRO0625; DNABPROTEIN.  
 DR PROSITE; PS00636; DNAB\_1; 1.  
 DR PROSITE; PS50076; DNAB\_2; 1.  
 KW Chapterone; Heat shock.  
 FT DOMAIN 1 70 J-DOMAIN.  
 SQ SEQUENCE 337 AA; 37806 MW; C7A9C613F73BCDAC CRC64;  
 Query Match 6.7%; Score 192; DB 1; Length 337;  
 Best Local Similarity 48.0%; Pred. No. 1.9e-06;  
 Matches 36; Conservative 16; Mismatches 19; Indels 4; Gaps 1;  
 QY 13 EDYSLNVRREASSEELKAAYRRLCMLYHPDKHDPDKSQARLFWLHQAAYVLS 72  
 3 KDYYCTLIGEKGSDDIDKAYRKALIKHPDKNSP-----QAEKFEVAEAYEVLSP 58  
 Db 73 QTRAIYDIYKRGLE 87  
 QY 59 KKEIREIYDYGEGRLK 73  
 RESULT 14  
 DJB5 HUMAN STANDARD; PRT; 348 AA.  
 ID DJB5 HUMAN 075953; O8TDR7;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE DnaJ homolog subfamily B member 5 (Heat shock protein Hsp40-3) (Heat shock protein cognate 40) (Hsc40) (Hsp40-2).  
 GN DNABJ5 OR HSC40.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OC NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Colon;  
 RX MEDLINE=20035738; PubMed=10570961;  
 RA Chen M.-S., Roti J.R., Laszlo A.;  
 RT Hsc40, a new member of the hsp40 family, exhibits similar expression profile to that of hsc70 in mammalian cells.;  
 RL Gene 238:333-341(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Fu Q., Yu L., Yue P., Zhou Y., Jiang J.X., Zhao S.Y.;  
 RT "Cloning and expression of a new human cDNA homology to human heat-

```

RT shock protein 40 mRNA."
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
CC -1- INDUCTION: Expressed under normal conditions, its expression can
CC further be increased after various stress treatments.
CC -1- SIMILARITY: Contains 1 J domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF088982; AAC35860.1; -.
DR EMBL; AF087870; AAM10498.1; -.
DR HSSP; P25685; 1HDJ.
DR Genew; HGNC:14887; DNABJ5.
DR GO; GO:0003773; F:heat shock protein activity; TAS.
DR GO; GO:0006950; P:response to stress; TAS.
DR InterPro; IPR002939; DnaJ_C.
DR InterPro; IPR001623; DnaJ_N.
DR InterPro; IPR008971; HSP40_DnaJ_pep.
DR InterPro; IPR003095; Hep_DnaJ.
DR Pfam; PF00226; DnaJ_1.
DR Pfam; PF01556; DnaJ_C_1.
DR PRINTS; PR00625; DNABPROTEIN.
DR PROSITE; PS00636; DnaJ_1; 1.
DR PROSITE; PS50076; DnaJ_2; 1.
DR Chaperone.
DR DOMAIN.
FT DOMAIN 4 68 MKITRRR -> IEDHKAS (IN REF. 2).
FT CONFLICT 190 196 TP -> HL (IN REF. 2).
FT CONFLICT 234 235
SQ SEQUENCE 348 AA; 39133 MW; DC9FG45DE4FD8CFC CRC64;

Query Match 6.7%; Score 192; DB 1; Length 348;
Best Local Similarity 43.6%; Pred. No. 2e-06;
Matches 34; Conservative 19; Mismatches 21; Indels 4; Gaps 1;

QY 13 EDYVSLINVRREASEELKAYRRLCNLYHDKRDELNSQAEFLNLYHQAYEVLSDP 72
DQ 3 KDYKIKIGIPSGANDEIKKAYRGMALKYHDKKEP-----NAEKEFKETAEAYDVLSDP 58
DB 73 QTRAIVDIYGRKGLUEMG 90
DB 59 KKGGLYDYGEGEGKLTG3 76

RESULT 15
DNU1_DROME STANDARD; PRT; 334 AA.
AC Q24133; Q9VRP0;
DT 01-NOV-1997 (Rel. 35, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DN DnaJ protein homolog 1 (DROJ1).
GN DnaJ-1 OR DROJ1 OR CG10578.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
ON
RN [1]
RP SEQUENCE FROM N.A.
RA Lee J.Y., Palter K.B.;
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

```

```

RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA April J.F., Agayanti A., An H.-O., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotier P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fouts R., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laeko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA Liu X., Matrei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Mostrelti A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun B.,
RA Svirskas R., Tector C., Turner R., Venter E.H., Wang A.H., Wang X.,
RA Wang Z.-Y., Weasman D.A., Weinstein G.M., Weisenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong P.N., Zhong W., Zhou X., Zhu S., Zhu H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley; TISSUE=Embryo;
RX MEDLINE=22426066; PubMed=12537369;
RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,
RA George R.A., Guarin H., Krommiller B., Pacle J.M., Park S., Wan K.H.,
RA Rubin G.M., Celisner S.E.;
RT "A Drosophila full-length cDNA resource."
RL Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- INDUCTION: By heat shock.
CC -1- SIMILARITY: Contains 1 J domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U34904; AAC23584.1; -.
DR EMBL; AB003565; AAP50753.1; -.
DR EMBL; AY058788; AAL14017.1; -.
DR HSSP; P25685; 1HDJ.
DR Pfam; P25685; 1HDJ.
DR FlyBase; FBgn0015657; DnaJ-1.
DR InterPro; IPR002939; DnaJ_C.
DR InterPro; IPR001623; DnaJ_N.
DR InterPro; IPR008971; HSP40_DnaJ_pep.
DR InterPro; IPR003095; Hep_DnaJ.
DR Pfam; PF00226; DnaJ_1.
DR Pfam; PF01556; DnaJ_C_1.
DR PRINTS; PR00625; DNABPROTEIN.
DR SMART; SM00271; DnaJ_1.
DR PROSITE; PS00636; DnaJ_1; 1.
DR PROSITE; PS50076; DnaJ_2; 1.
DR Chaperone; Heat shock.
FT DOMAIN 4 68 J-DOMAIN.

```

FT	CONFLICT	261	261	G -> E (IN REF. 1).
SQ	SEQUENCE	334	AA; 37028	MM; 60542ABFD47A5689
				CRC64;

SQ SEQUENCE 334 AA; 37028 MW; 60542ABFD47A5689 CRC64;

Query Match	6.78;	Score 191;	DB 1;	Length 334;
Best Local Similarity	48.08;	Prod No 3	20-05-	

Matches 36; Conservative 17; Mismatches 18; Indels 4; Gaps 1;

Matches 36; Conservative 17; Mismatches 18; Indels 4; Gaps 1;

```
QY      13 EDYSLINVRKRASSSEELKAAVRLCMLYHPDKRDPELKSQAELFLNVHQAYEVLSDP  72
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      3 KDFYKILGLERRASDDEIKAAVRLATLKVHPDNKSP-----QAEERFKEIAEAYEVLSDK  58
```

Db 3 KDFYKILGLERKASDDEIKAYRKLALKYHPDKNSP---QAERFKEIAAEYEVLSDK 58

QY 73 QTRAIYDIYKRGLE 87

Db 59 KKRDI FDNYGEGDLK 73

Search completed: July 13, 2004, 08:41:07  
Job time : 19 secs

Job time : 19 secs





XX	Modified-site	362	/note= "potential phosphorylation site"
FT	Modified-site	448	/note= "potential phosphorylation site"
FT	Modified-site	477	/note= "potential phosphorylation site"
FT	Modified-site	480	/note= "potential phosphorylation site"
FT	Modified-site	552	/note= "potential phosphorylation site"
XX	WO200017358-A2.		
PN	30-MAR-2000.		
PD	22-SEP-1999;		
PF	99WO-US022027.		
XX	22-SEP-1998;		
PR	98US-00150642.		
PR	22-SEP-1998;		
PR	98US-0172221P.		
PR	19-JAN-1999;		
PR	99US-00233291.		
PR	19-APR-1999;		
PR	99US-00294698.		
XX	19-APR-1999;		
XX	99US-0172232P.		
PA	(INCY-) INCYTE PHARM INC.		
PI	Tang YT, Hillman JL, Yue H, Patterson C, Baughn MR, Batra S;		
XX	WPI; 2000-283583/24.		
DR	N-PSDB; AAA08568.		
XX	New purified polypeptides and polynucleotides encoding human chaperone proteins, useful for diagnosing, treating and preventing disorders associated with the expression human chaperone proteins.		
PT	Claim 1; Page 75-76; 89pp; English.		
PS	AAA08563-68 encode human chaperone proteins 1-6 (HCHP-1 to HCHP-6) respectively. The sequences can be used to treat and prevent disorders associated with altered expression or activity of HCHP comprising administering a composition comprising the polypeptide or an antagonist to a patient (claimed). The human chaperone proteins are also useful for the diagnosis, treatment or prevention of neurodegenerative, metabolic, developmental, autoimmune/inflammatory disorders and cell proliferative disorders including cancer		
CC	Sequence 559 AA;		

	Query Match	100.0%; Score 2857; DB 3; Length 559;
	Best Local Similarity	100.0%; Pred. No. 1.4e-263;
	Matches 559; Conservative	0; Mismatches 0; Indels 0; Gaps 0.
QY	1	MATAISEELNEDDYSLILNVAREASSEELKAAAYRLCMLYHPDGHDPBELKSQAERLFN 60
Db	1	MATLASEELNEDDYSLILNVRRRESSSELKAAAYRLCMLYHPDGHDPBELKSQAERLFN 60
QY	61	LVAQAEVLSDPQTPRAIYDIYGKRGLEMEGWEVERRRTPAIREEFERLQREBERRLQ 120
Db	61	LVAQAEVLSDPQTPRAIYDIYGKRGLEMEGWEVERRRTPAIREEFERLQREBERRLQ 120
QY	121	QRTNPKGITSGVDATDLFDRIDEYEDEVSSGFPOIINMKHISQSIEAPLTATDAI 180
Db	121	QRTNPKGITSGVDATDLFDRIDEYEDEVSSGFPOIINMKHISQSIEAPLTATDAI 180
QY	181	SGSLSTONGNGGGSINFALRRVTSAKMGGELEFGAGDIQGPFGLKLFRNLTPRCFTYTN 240
Db	181	SGSLSTONGNGGGSINFALRRVTSAKMGGELEFGAGDIQGPFGLKLFRNLTPRCFTYTN 240
QY	241	CALQFSRGRIPGLTTVLARNLDKNTVGYLQRMWIGSMMNTSYIRDPTKSHFTVALDG 300
Db	241	CALQFSRGRIPGLTTVLARNLDKNTVGYLQRMWIGSMMNTSYIRDPTKSHFTVALDG 300
QY	301	IHPSPALISYOHKKFDQDDQTRVKGSLKAGPGTVVEYAERRKISHSVTLGAASGVGPG 360

Db	301	IPHSFALLISYQHKFDDDOQTRKXSLKAGFFGTVEVGAERKXISRHSLVLAASVGPQG	360
Qy	361	VSLSKTKLNASQTYTFFPHILTDQLPSAMPATYAGPLVYTFAMRLIIKPYLPAQKEKL	420
Db	361	VSLSKTKLNASQTYTFFPHILTDQLPSAMPATYAGPLVYTFAMRLIIKPYLPAQKEKL	420
Qy	421	EKQRESAADTVLOKQOEBSAARLMQESVRRILIEAESRNGLLIVNMYKPFNDSSRS	480
Db	421	EKQRESAADTVLOKQOEBSAARLMQESVRRILIEAESRNGLLIVNMYKPFNDSSRS	480
Qy	481	EKKVVDVTVPLQCLVKQSKLILTEASKAGPGFYDPCVGEENKLYLQFPGVLQVWV	540
Db	481	EKKVVDVTVPLQCLVKQSKLILTEASKAGPGFYDPCVGEENKLYLQFPGVLQVWV	540
Qy	541	LDSEALRIPKQSHRIDTDG	559
Db	541	LDSEALRIPKQSHRIDTDG	559
RESULT 2			
ID	ABG79329	standard; protein; 559 AA.	
XX	ABG79329;		
DT	15-NOV-2002	(first entry)	
XX			
DE		Human foetal cell selective cDNA encoded protein #28.	
XX			
KM		Human; foetal liver myeloid cell; erythroblast; foetal abnormality;	
KM		maternal blood; differential display; chromosomal abnormality;	
KM		single gene disorder; aneuploidy; nucleotide triplet expansion disorder;	
KM		trisomy 13; trisomy 21; Klinefelter syndrome; spina bifida;	
KM		sickle cell anemia; thalassemia; Marfan syndrome;	
KM		sickle muscular dystrophy; cystic fibrosis; Fragile X-syndrome;	
KM		Friedreich's ataxia; myotonic dystrophy; Huntington's disease;	
KM		sex determination.	
OS		Homo sapiens.	
XX			
PN	W0200255985-A2.		
XX			
PD	18-JUL-2002.		
XX			
PF	01-NOV-2001; 2001WO-US045340.		
XX			
PR	15-NOV-2000; 2000US-0248882P.		
XX			
PA	(HOFF ) ROCHE DIAGNOSTICS CORP.		
XX			
P1	Schueler PA, Xu H, Foltz L, Wu X, Sha Y, Nagy A, Mahoney WC;		
XX			
DR	WPI; 2002-619108/66.		
XX			
DR	N-PSDB; ABS64608.		
XX			
PT			
PT	Detecting fetal cells in maternal blood sample, useful for diagnosing an		
PT	abnormality in a fetal cell, involves using specific nucleic acid probes		
XX	that hybridize to fetal cell associated RNAs.		
PS	Claim 116; Page 204-205; 215pp; English.		
XX			
CC	The invention relates to detecting foetal cells in maternal blood sample,		
CC	involving contacting sample with a first probe or performing expression		
CC	analysis on RNA/cDNA obtained from foetal liver myeloid cells relative to		
CC	CC that obtained from mature cells to identify RNA/cDNA species useful as		
CC	probe, contacting sample with the probe and identifying if the sample		
CC	comprises a cell that comprises mRNA that hybridises to the probe. The		
CC	probes are identified by differential display analysis using mature liver		
CC	cells and foetal liver myeloid cells of less than 22 weeks of gestation.		
CC	Also included are: (1) The probe sequences (appearing as ABS64586-		
CC	ABS64618), an isolated nucleic acid molecule having a sequence which is		
CC	at least 90% identical to the probe or its complement, or having a		
CC	nucleotide sequence identical to at least 15 consecutive nucleotide		

CC residues of the probe, the encoded proteins from the probe or encoding a  
CC naturally occurring allelic variant or fragment; (2) a non-mammalian host  
CC cell containing the probe and (3) an antibody which selectively binds to  
CC the probe encoded protein. The method is useful for detecting a foetal  
CC cell (such as erythroblast or trophoblast) in a maternal blood sample,  
CC which is useful for diagnosing a chromosomal abnormality, single gene  
CC disorder or nucleotide triplet expansion in the gene, in a foetal cell.  
CC For example an aneuploidy (trisomy 13, trisomy 21, or Klinefelter  
CC syndrome), spina bifida, sickle cell anaemia, a thalassaemia, Marfan  
CC syndrome, Duchenne muscular dystrophy, cystic fibrosis, Fragile X-  
CC syndrome, Friedreich's ataxia, myotonic dystrophy or Huntington's  
CC disease. The probes are also useful for determination of the sex of a  
CC foetus and for detecting the presence of multiple foetuses at early  
CC stages of pregnancy. The present sequence is encoded by a foetal cell  
CC specific probe of the invention

XX Sequence 559 AA;

Query Match 100.0%; Score 2857; DB 5; Length 559;

Best Local Similarity 100.0%; Pred. No. 1.4e-263; Mismatches 0; Gaps 0;

Matches 559; Conservative 0; Indels 0; Gaps 0;

QY 1 MATALSEBELDNEEDYSLNVRREASSELKAAAYRRLCMLYHPDKHDPBELKSOAERLFN 60

DB 1 MATALSEBELDNEEDYSLNVRREASSELKAAAYRRLCMLYHPDKHDPBELKSOAERLFN 60

QY 61 LVHQAAYEVLSDPQTRAIYDIYGRKGLMEGWEEVERRTPAIRREPERLQREBERRLQ 120

DB 61 LVHQAAYEVLSDPQTRAIYDIYGRKGLMEGWEEVERRTPAIRREPERLQREBERRLQ 120

QY 121 QRTNPKGTISVGVADATLFDRIYDEYEDVSGSSFPQIEINMKHISQSIETAPLTATDAIL 180

DB 121 QRTNPKGTISVGVADATLFDRIYDEYEDVSGSSFPQIEINMKHISQSIETAPLTATDAIL 180

QY 181 SGSLSTONGNGGSGINFALRRVTSAGKWELEFGAGDLQGPLFGKLFRNLTPRCFVTYN 240

DB 181 SGSLSTONGNGGSGINFALRRVTSAGKWELEFGAGDLQGPLFGKLFRNLTPRCFVTYN 240

QY 241 CALQFSSRGIRPGLTTLVARNLDKRTVGYLQWRKCIQSANMTSYRDTKTSHTFTVALQLG 300

DB 241 CALQFSSRGIRPGLTTLVARNLDKRTVGYLQWRKCIQSANMTSYRDTKTSHTFTVALQLG 300

QY 301 IPHSFALISYOHKFODDQTRVKSGLKAGFGTVVEYGAERKISRHSVLAAGAVGVPOG 360

DB 301 IPHSFALISYOHKFODDQTRVKSGLKAGFGTVVEYGAERKISRHSVLAAGAVGVPOG 360

QY 361 VSLKVKLNRAASQTYFFPIHLTDQLPSAMFYATVGPVVVYPMHRLIKPYLRAQKEXEL 420

DB 361 VSLKVKLNRAASQTYFFPIHLTDQLPSAMFYATVGPVVVYPMHRLIKPYLRAQKEXEL 420

QY 421 EKORBSAATDVLOKKQEAESAVALMOESVRLIIEAESRMGLIYNAMYGFVNDKSRKS 480

DB 421 EKORBSAATDVLOKKQEAESAVALMOESVRLIIEAESRMGLIYNAMYGFVNDKSRKS 480

QY 481 EKVVIVDTVPLQCLVKOSKILLTEASKAGLPGFYDPVGVGEKNLKVLYORGVLAHQVMV 540

DB 481 EKVVIVDTVPLQCLVKOSKILLTEASKAGLPGFYDPVGVGEKNLKVLYORGVLAHQVMV 540

QY 541 LDSEALRIPOKSHRIDTDG 559

DB 541 LDSEALRIPOKSHRIDTDG 559

RESULT 3

AAB93115 ID AAB93115 standard; protein; 559 AA.

XX AAB93115;

XX 26-JUN-2001 (first entry)

XX Human protein sequence SEQ ID NO:11987.

KM Human; primer; detection; diagnosis; antisense therapy; gene therapy.

XX Homo sapiens.

OS Homo sapiens.

PN EPI074617-A2.

XX EPI074617-A2.

PD 07-FEB-2001.

XX 07-FEB-2001.

PF 28-JUL-2000; 2000EP-00116126.

XX 28-JUL-2000; 2000EP-00116126.

PR 29-JUL-1999; 99JP-00248036.

XX 29-JUL-1999; 99JP-00248036.

PR 27-AUG-1999; 99JP-00300253.

XX 27-AUG-1999; 99JP-00300253.

PR 11-JAN-2000; 2000JP-00118776.

XX 11-JAN-2000; 2000JP-00118776.

PR 02-MAY-2000; 2000JP-00183767.

XX 02-MAY-2000; 2000JP-00183767.

PR 09-JUN-2000; 2000JP-00241899.

XX 09-JUN-2000; 2000JP-00241899.

XX (HELI-) HELIX RES INST.

XX (HELI-) HELIX RES INST.

PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

PI Ishii S, Sugiyama T, Wakamatsu A, Negai K, Otsuki T;

XX Ishii S, Sugiyama T, Wakamatsu A, Negai K, Otsuki T;

DR WPI; 2001-318749/34.

XX WPI; 2001-318749/34.

PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-

XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-

PT length cDNAs defined in the specification, and for the detection and/or

XX length cDNAs defined in the specification, and for the detection and/or

PT diagnosis of the abnormality of the proteins encoded by the full-length

XX diagnosis of the abnormality of the proteins encoded by the full-length

PT cDNAs.

XX cDNAs.

PS Claim 8; SEQ ID NO 11987; 2537pp + Sequence Listing; English.

XX Claim 8; SEQ ID NO 11987; 2537pp + Sequence Listing; English.

CC The present invention describes primer sets for synthesizing 5602 full-

XX The present invention describes primer sets for synthesizing 5602 full-

CC length cDNAs defined in the specification. Where a primer set comprises:

XX length cDNAs defined in the specification. Where a primer set comprises:

CC (a) an oligo-dT primer and an oligonucleotide complementary to the

XX (a) an oligo-dT primer and an oligonucleotide complementary to the

CC complementary strand of a polynucleotide which comprises one of the 5602

XX complementary strand of a polynucleotide which comprises one of the 5602

CC nucleotide sequences defined in the specification, where the

XX nucleotide sequences defined in the specification, where the

CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination

XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination

CC of an oligonucleotide comprising a sequence complementary to the

XX of an oligonucleotide comprising a sequence complementary to the

Db 181 SGLSTONGNGGGSINFALRRVTSKAGWGELEFGAGDLOGPLFGKLFRNLTPRCFVTNN 240  
 QY 241 CALQFSSRGIRPGLTTVLAARNLDKNTVGYLQWRMGIOSAMNTSIVRDTKTSHTFVALQIG 300  
 Db 241 CALQFSSRGIRPGLTTVLAARNLDKNTVGYLQWRMGIOSAMNTSIVRDTKTSHTFVALQIG 300  
 QY 301 IPHSFALISYQHKFODDQTRVKGSLKAGFGTVVEYGABRKISRHSVLGAASVGPQG 360  
 Db 301 IPHSFALISYQHKFODDQTRVKGSLKAGFGTVVEYGABRKISRHSVLGAASVGPQG 360  
 QY 361 VSLKVKLNRASTQYFFPIHITDQLLPSAMFYATVGPVYTFAMHRLIKYLLAOKREKL 420  
 Db 361 VSLKVKLNRASTQYFFPIHITDQLLPSAMFYATVGPVYTFAMHRLIKYLLAOKREKL 420  
 QY 421 EKQRESAATVLOKQOAGSAVRLMQESVRIIEAESRMGLIIVNAMYGFVNDKSRKS 480  
 Db 421 EKQRESAATVLOKQOAGSAVRLMQESVRIIEAESRMGLIIVNAMYGFVNDKSRKS 480  
 QY 481 EKXKVIDVTVPLQCLVDSKLIILTEASKAGLPGFYDPCEGEKRLKVLXOFRGVLHQVMV 540  
 Db 481 EKXKVIDVTVPLQCLVDSKLIILTEASKAGLPGFYDPCEGEKRLKVLXOFRGVLHQVMV 540  
 QY 541 LDSEALRIPKQSHRIDTDG 559  
 Db 541 LDSEALRIPKQSHRIDTDG 559

## RESULT 4

ABG79328 standard; protein; 336 AA.

ABG79328;

15-NOV-2002 (first entry)

Human foetal cell selective cDNA encoded protein #27.

XX Human; foetal liver myeloid cell; erythroblast; foetal abnormality;  
 XX maternal blood; differential display; chromosomal abnormality;  
 XX single gene disorder; aneuploidy; nucleotide triplet expansion disorder;  
 XX trisomy 13; trisomy 21; Klinefelter syndrome; spina bifida;  
 XX sickle cell anaemia; thalassemia; Marfan syndrome;  
 XX Duchenne muscular dystrophy; cystic fibrosis; Fragile X-syndrome;  
 XX Friedreich's ataxia; myotonic dystrophy; Huntington's disease;  
 XX sex determination.

OS Homo sapiens.

PN W0200255985-A2.

PD 18-JUL-2002.

PF 01-NOV-2001; 2001WO-US045340.

PR 15-NOV-2000; 2000US-0248882P.

XX (HOF) ROCHE DIAGNOSTICS CORP.

PI Schueler PA, Xu H, Folz L, Wu X, Sha Y, Nagy A, Mahoney WC;

DR WPI, 2002-619108/66.

DR N-PSDB; ABS64607.

PT Detecting fetal cells in maternal blood sample, useful for diagnosing an  
 PT abnormality in a fetal cell, involves using specific nucleic acid probes  
 PT that hybridize to fetal cell associated RNAs.

XX Claim 116; Page 202-203; 215pp; English.

CC The invention relates to detecting foetal cells in maternal blood sample,  
 CC involving contacting sample with a first probe or performing expression  
 CC analysis on RNA/cDNA obtained from foetal liver myeloid cells relative to  
 CC that obtained from mature cells to identify RNA/cDNA species useful as

CC probe, contacting sample with the probe and identifying if the sample  
 CC comprises a cell that comprises mRNA that hybridises to the probe. The  
 CC probes are identified by differential display analysis using mature liver  
 CC cells and foetal liver myeloid cells of less than 22 weeks of gestation.  
 CC Also included are: (1) The probe sequences (appearing as ABS64586-  
 CC ABS64618), an isolated nucleic acid molecule having a sequence which is  
 CC at least 90% identical to the probe or its complement, or having a  
 CC nucleotide sequence identical to at least 15 consecutive nucleotide  
 CC residues of the probe, the encoded proteins from the probe or encoding a  
 CC naturally occurring allelic variant or fragment; (2) a non-mammalian host  
 CC cell containing the probe and (3) an antibody which selectively binds to  
 CC the probe encoded protein. The method is useful for detecting a foetal  
 CC cell (such as erythroblast or trophoblast) in a maternal blood sample,  
 CC which is useful for diagnosing a chromosomal abnormality, single gene  
 CC disorder or nucleotide triplet expansion in the gene, in a foetal cell.  
 CC For example an aneuploidy (trisomy 13, trisomy 21, or Klinefelter  
 CC syndrome), spina bifida, sickle cell anaemia, a thalassemia, Marfan  
 CC syndrome, Duchenne muscular dystrophy, cystic fibrosis, Fragile X-  
 CC syndrome, Friedreich's ataxia, myotonic dystrophy or Huntington's  
 CC disease. The probes are also useful for determination of the sex of a  
 CC foetus and for detecting the presence of multiple foetuses at early  
 CC stages of pregnancy. The present sequence is encoded by a foetal cell  
 CC specific probe of the invention

Sequence 336 AA;

Query Match 60.3%; Score 1723; DB 5; Length 336;

Best Local Similarity 99.1%; Pred. No. 1.5e-155; Matches 333; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MATALSEELDNEDYSLNVRREASSEELKAAYRRLCMLYHPDKHDPBELKQABRLFN 60  
 Db 1 MATALSEELDNEDYSLNVRREASSEELKAAYRRLCMLYHPDKHDPBELKQABRLFN 60  
 QY 61 LVHQAAYEVSIDPQTRAIYDIYGRGLIEMGEWVERRRTPAIRREFFERLQREBERRLQ 120  
 Db 61 LVHQAAYEVSIDPQTRAIYDIYGRGLIEMGEWVERRRTPAIRREFFERLQREBERRLQ 120  
 QY 121 QRTNPGRITSVGVNATDLDREYDEREDVSGSPQIEINKKHISOSTRAPLTATDTALL 180  
 Db 121 QRTNPGRITSVGVNATDLDREYDEREDVSGSPQIEINKKHISOSTRAPLTATDTALL 180  
 QY 181 SGLSTONGNGGGSINFALRRVTSKAGWGELEFGAGDLOGPLFGKLFRNLTPRCFVTNN 240  
 Db 181 SGLSTONGNGGGSINFALRRVTSKAGWGELEFGAGDLOGPLFGKLFRNLTPRCFVTNN 240  
 QY 241 CALQFSSRGIRPGLTTVLAARNLDKNTVGYLQWRMGIOSAMNTSIVRDTKTSHTFVALQIG 300  
 Db 241 CALQFSSRGIRPGLTTVLAARNLDKNTVGYLQWRMGIOSAMNTSIVRDTKTSHTFVALQIG 300  
 QY 301 IPHSFALISYQHKFODDQTRVKGSLKAGFGTVVE 336  
 Db 301 IPHSFALISYQHKFODDQTRVKGSLKAGFGTVVE 336

## RESULT 5

AAM39772 standard; protein; 356 AA.

AAM39772;

22-OCT-2001 (first entry)

Human polypeptide SEQ ID NO 2917.

XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
 XX peripheral nervous system; neuropathy; central nervous system; CNS;  
 XX Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
 XX amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
 XX chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
 XX leukaemia.

OS Homo sapiens.





cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (AB161617-AB130511), expressed DNA sequences (AB101840-AB161175) and the encoded proteins (AB57737-7-AB572072). The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [http://www.wipo.int/pub/published\\_pat\\_sequences](http://www.wipo.int/pub/published_pat_sequences)

SQ Sequence 521 AA;

Query Match	39.5%;	Score 1128.5;	DB 4;	Length 521;
Best Local Similarity	44.2%;	Pred. No. 2e-98;		
Matches 233;	Conservative 112;	Mismatches 169;	Indels 13;	Gaps 6;

Qy	6	SEELJNEDYSLNLRREKSSSEELAAARRICMLYHPKHPDEPKSAQARLFNLVYQA	65
Db	8	SDAELD-ENYTFPLN.PROATAQIINTAARKQSRMHDPKHDLPDCKMAEIMFRTRGA	66
Qy	66	YELVSDPQPAIYDIYKSGKLEMEGHEVYERRRTPAETREERLQREERRLQORTNP	122
Db	67	YELVSDPQPAIYDVGEGKLRTEGEBILHRTKTPDELTREERYERLAQAAARRRLQORTNP	126
Qy	126	KGTISVGDATDIDFRDEREDVDSSGSPFOIEINMHISQSIEMAPLATATPALSGLS	180
Db	127	RGNITINVMNTEIFAAYDD-----SEMPHEIGSMISAQIETAPITRDOMIMSGULY	179
Qy	186	TONGGGSINPALARVTSACKGMBELBPAGDLOBPGLKLPNLRTRCFVTNICALQF	245
Db	180	SSNGNSGGGVIAGRRLN-KGWIELCAQAGN--GFLGLKGKRTLISOQLTLANGTINLF	233
Qy	246	SSRGIRGLTTLAARNLDKRYTVGYLQWRNGIOSAMNTSIVRDTKTSHTFVALQIGIPSF	305
Db	237	RDCGVIPALFSTLAVALQDGRHTMSGLTLNAGSSMSHTQDHSKETYSLSSSVIGTPIHY	296
Qy	306	ALISYOHKFODDQTVRKVSKLGAFFGVYEVYGAERKISRHSVLGAASVGVPOGVSILKV	365
Db	297	FGLSYTERKMA-ENELKTLAKVGTGFGEYGEVEKRSKYSAPATATISIGVSGVILKF	355
Qy	366	KLRASQTYEFPRLHTDOLLPASMFATYGLPVYFAMHRLIKRYLALQREKELKQRE	422
Db	356	KLRNSQSYVFPRLHSDDEIVPAVVPASVATPVLAWEFIKRYTDMDEMERKNTIEVERKJR	411
Qy	426	SAATDVLOKKOAEASAVRLMOESVRRIIAEBSRMGLITVNAWYKFPVNDKSRKEKYV	485
Db	416	ONEQRISAKRHEASAAVHLMQATYNNIMTEELARNGLIATRAVYUCTIAG-GRQKRPQS	474
Qy	486	IDTVTPLOGLYKDSKILTEBASKAGLPGYVDPCEBEKTKLYLQFR	532
Db	475	LDVTVALQCVWKNQTLQHDSSSKSDLPFGYDPIGIDKTLREIYYQ	521

## RESULT 9

ID ABG03199 standard; protein; 344 AA.

AC ABG03199

DT 13-FEB-2002 (first entry)

**DE** Novel human diagnostic protein #3190.

**KW** Human; chromosome mapping; gene mapping; gene therapy; forensic;

2 XX

XX XX

PN MO200175067-A2.

11-OCT-2001  
PD  
yy

30-MAR-2001; 2001WO-US008631.

PR	31-MAR-2000; 2000US-00540217.
PR	23-AUG-2000; 2000US-00649167.
...	

PA (HYSE-) HYSEQ INC.

PI Drymanac RT, Liu C

DR WPI; 2001-639362/73.

2  
 3  
 4  
 5  
 6  
 7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525  
 526

PT diagnostics, fore

biodiversity.

PS Claim 20; SEQ ID NO 33558; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and polypeptide (II)

CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
CC and in recombinant production of (II). The polynucleotides are also used  
CC in diagnostics as expressed sequence tags for identifying expressed  
CC genes. (I) is useful in gene therapy techniques to restore normal  
CC activity of (II) or to treat disease states involving (II). (II) is  
CC polypeptide for generating antibodies against it, detecting or quantitating a  
CC polypeptide in tissue, as molecular weight markers and as a food  
CC supplement. (II) and its binding partners are useful in medical imaging  
CC of sites expressing (II). (I) and (II) are useful for treating disorders  
CC involving aberrant protein expression or biological activity. The  
CC polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AB000010-AB030377 represent novel human diagnostic  
CC amino acid sequences of the invention. Note: The sequence data for this  
CC patent did not appear in the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC [http://wipo.int/pub/published\\_pat\\_sequences](http://wipo.int/pub/published_pat_sequences)

SQ Sequence 344 AA;

Query Match	39.3%	Score 1124	DB 4	Length 344
Best Local Similarity	84.1%	Pred. No. 2.7e-98		
Matches 232; Conservative	3;	Mismatches 21;	Indels 20;	Gaps 4;

QY	1	MAALASBEELBDEDDYSSILNIVREASSSEELKAAYRLOMLVHPDGGRPELK----	LMK	55
Db	61	LVHQAQVEVLSDBQTRAIY-----DIQKQGLMEQNEVVERRRTPAIEIRREERLOREKEE		116
QY	56	CL-----VTPKQCPSPMIYMGREDWMXKQRLWIKGCBP-----LEIRREERLIREEE		104
Db	117	RRLQORTNPKGTSVGVDAITDLFDRYDEEYEDVESSSPQJIEINMGHSQSIEAPLPTATD		174
QY	105	RRLQGRTNPKGTTISVGVDAITDLFDRYDEEYEDVESSSPQJIEINMGHSQSIEAPLPTATD		164
Db	177	TALISSELSITONGNGGGSINPALREVTSAKMGSELPEPAGLOGLPGLKTLPRVLTPLPCF		234
QY	165	TALISSELSITONGNGGGSINPALRRVTSKMGSELPEPAGLOGLPGLKTLPRVLTPLPCF		224
Db	237	VTTNCAQFSSRGIRPGLTTVLAARNLDKXNVGYLQW		272
QY	225	VTTNCAQFSSRGIRPGLTTVLAARNLDKXNVGYLQW		260

## RESULT 10

ID AU17215 standard; protein; 159 AA.

AC AAU17215;

DT 07-NOV-2001 (first entry)

XX DE Novel signal transduction pathway protein, Seg ID 780.  
XX KW Neuroprotective; cytosolic; dermatological; immunosuppressive; tumour;  
KW Antinflammatory; anti-HIV; antibacterial; antineoplastic; cancer;  
KW Immune system disorder; rheumatoid arthritis; inflammatory condition;  
KW Organ transplant rejection; infection; hepatitis C; blood disorder;  
KW sickle cell anaemia; hyperproliferative disorder; Gaucher's disease;  
KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;  
KW chromosomal abnormality; Down syndrome; ischaemia; renal disorder;  
KW cardiovascular; respiratory; wound healing; endocrine; Addison's disease;  
KW reproductive system; gastrointestinal; liver disorder; AIDS;  
KW acquired immune deficiency syndrome.  
XX OS Homo sapiens.  
XX PN WO200154733-A1.  
XX PD 02-AUG-2001.  
XX 17-JAN-2001; 2001WO-US001312.  
XX PF 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 14-AUG-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226686P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.

PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235634P.  
PR 27-SEP-2000; 2000US-0235635P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249246P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.





PI Rosen CA, Ruben SM, Barash SC;  
XX WPI; 2003-719985/68.  
DR N-PSDB; ADB93300.  
XX  
PT New isolated polypeptide useful for diagnosing and treating  
PT immunosuppressive conditions such as autoimmune disease and Parkinson's  
PT disease.  
PS  
PS Claim 11; SEQ ID NO 780; 345pp; English.  
XX  
XX The invention relates to an isolated polypeptide. The polypeptide is  
CC useful for diagnosing a pathological condition or a susceptibility to a  
CC pathological condition in a subject, by determining the presence or  
CC amount of expression of the polypeptide in a biological sample and  
CC diagnosing a pathological condition or a susceptibility to a pathological  
CC condition based on the presence or amount of expression of the  
CC polypeptide. The polypeptide is also useful for identifying a binding  
CC partner to the polypeptide, which involves contacting the polypeptide  
CC with a binding partner and determining whether the binding partner  
CC effects an activity of the polypeptide. The polypeptide or the nucleic  
CC acid encoding the polypeptide is useful for preventing, treating, or  
CC ameliorating a medical condition, which involves administering the  
CC polypeptide or the nucleic acid to a mammalian subject. The nucleic acid  
CC is useful for diagnosing a pathological condition or a susceptibility to  
CC a pathological condition in a subject, which involves determining the  
CC presence or absence of a mutation in the nucleic acid, and diagnosing a  
CC pathological condition or susceptibility to a pathological condition  
CC based on the presence or absence of the mutation. The polypeptide, the  
CC nucleic acid and an antibody to the polypeptide are useful for treating  
CC autoimmune disease, Parkinson's disease, sickle cell disease, Crohn's  
CC disease, atherosclerosis, haemophilia, thrombocytopenia. The polypeptide,  
CC the nucleic acid and the antibody are useful as immunosuppressive agents,  
CC as adjuvants to enhance immune responses, and as agents to induce higher  
CC affinity antibodies and increase serum immunoglobulin concentrations. The  
CC present sequence represents the amino acid sequence of a novel human  
CC protein. Note: The sequence data for this patent did not form part of the  
CC printed specification but was obtained in electronic format direct from  
CC USPTO at [seqdata.uspto.gov/sequence.html?docid=20020168711](http://seqdata.uspto.gov/sequence.html?docid=20020168711).  
XX  
XX Sequence 159 AA;  
SQ  
Query Match 26.3%; Score 750; DB 7; Length 159;  
Best Local Similarity 92.5%; Pred. No. 4,4e-63;  
Matches 147; Conservative 1; Mismatches 11; Indels 0; Gaps 0;  
QY 2 ATATSEELNEDYYSILNRRREASSSEELKAAARRLCMLYHPKRDPELKSQAERLFL 61  
DB 1 ATATSEELNEDYYSILNRRREASSSEELKAAARRLCMLYHPKRDPELKSQAERLFL 60  
QY 62 VHQAYEVLSDPQRAIDYIGKRGLEMEGMEVVERRTPTAIREEPERLQREERERLQ 121  
DB 61 VHQAYEVLSDPQRAIDYIGKRGLEMEGMEVVERRTPTAIREEPERLQREERERLQ 120  
QY 122 RINPKGTISGVVDATDLFDREYDEYEDVSGSSFPQIEIN 160  
DB 121 RINPKGTISGVVDATDLFDREYDXKLKWCPAVXFXPIEIN 159  
RESULT 12  
AAU17589  
ID AAU17589 standard; protein; 159 AA.  
XX  
XX AAU17589;  
AC  
DT 07-NOV-2001 (first entry)  
XX  
XX Novel signal transduction pathway protein, Seq ID 1154.  
DE  
XX  
XX Neuroprotective; cytostatic; dermatological; immunosuppressive; tumour;  
KW antiinflammatory; anti-HIV; antibacterial; antiinflammatory; cancer;  
KW immune system disorder; rheumatoid arthritis; inflammatory condition;  
KW organ transplant rejection; infection; hepatitis C; blood disorder;

KW sickle cell anaemia; hyperproliferative disorder; Gaucher's disease;  
KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;  
KW chromosomal abnormality; Down syndrome; ischaemia; renal disorder;  
KW cardiovascular; respiratory; wound healing; endocrine; Addison's disease;  
KW reproductive system; gastrointestinal; liver disorder; AIDS;  
KW acquired immune deficiency syndrome.  
XX  
XX Homo sapiens.  
OS  
PN WO200154733-A1.  
XX  
XX 02-AUG-2001.  
PD  
XX 17-JAN-2001; 2001WO-US001312.  
PF  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226868P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228624P.  
PR 01-SEP-2000; 2000US-0229387P.  
PR 01-SEP-2000; 2000US-0229387P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 05-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.

PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 29-SEP-2000; 2000US-0235835P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251866P.

PR 08-DEC-2000; 2000US-0251866P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
PI Rosen CA, Barash SC, Ruben SM;  
XX  
XX WPI; 2001-465460/50.  
DR N-PSDB; AAS27506.  
XX  
PT Novel polypeptides useful for diagnosing, treating, preventing and/or  
PT prognosing disorders related to the proteins, including cancers, immune  
PT disorders and neuronal disorders.  
XX  
XX Claim 1; SEQ ID NO 1154; 880pp; English.  
XX  
CC The invention relates to novel isolated polypeptides (I), and  
CC polynucleotides (II). (I), (II) and the antibody to (I) are useful for  
CC diagnosing, preventing and treating diseases including immune system  
CC disorders (e.g. congenital and acquired immunodeficiencies, autoimmune  
CC disorders (e.g. rheumatoid arthritis), inflammatory conditions, organ  
CC transplant rejections and graft versus host disease, infectious diseases  
CC (e.g. hepatitis C), bleeding disorders, haemoglobin abnormalities and  
CC other blood-related disorders (sickle cell anaemia), myeloproliferative  
CC disorders, primary haematopoietic disorders, hyperproliferative disorders  
CC (e.g. Gaucher's disease and cancer), neurodegenerative disorders (e.g.  
CC Alzheimer's disease, Parkinson's disease), chromosomal abnormalities  
CC (Down syndrome), ischaemic injury (e.g. stroke), renal disorders (e.g.  
CC glomerulonephritis), cardiovascular disorders (e.g. arrhythmia),  
CC respiratory disorders, dermatological disorders, in wound healing,  
CC epithelial cell proliferation, endocrine disorders (e.g. Addison's  
CC disease), reproductive system disorders, gastrointestinal disorder  
CC (inflammatory disorders), liver disorders (cirrhosis), as stimulators of  
CC B-cell responsiveness to pathogens, activators of T-cells, to induce  
CC higher affinity antibodies, and as a means to induce tumour proliferation  
CC in pathologies e.g. acquired immune deficiency syndrome (AIDS). AAU17059-  
CC AAU17683 represent novel signal transduction pathway protein, amino acid  
XX sequences of the invention

Query Match 25.7%; Score 733; DB 4; Length 159;  
Best Local Similarity 90.6%; Pred. No. 1.9e-61;  
Matches 144; Conservative 1; Mismatches 14; Indels 0; Gaps 0;

QY 2 ATATSEELNEDYSLANVRREASSSELKAAYRRLQMLVHPKHPPELKSQAERFLN 61  
Db 1 ATATSEELNEDYSLANVRREASSSELKAAYRRLQMLVHPKHPPELKSQAERFLN 60  
QY 62 VHQAYEVLSDPQTRAIYIGKRGLEMEGWEVVERRTTPAIREEFRLQREBERRLQ 121  
Db 61 VHQAYEVLSDPQTRAIYIGKRGLEMEGWEVVERRTTPAIREEFRLQREBERRLQ 120  
QY 122 RTNPKGTISGVNDATLDFRYDEBEYDVSSSPQQLIN 160  
Db 121 RTNPKGTISGVNDATLDFRYDXXLKKCPAVXFXIEIN 159

RESULT 13  
ADB94297 standard; protein; 159 AA.  
ID ADB94297;  
XX  
AC ADB94297;  
XX  
DT 04-DEC-2003 (first entry)  
XX  
DS Human novel protein #531.  
XX human, autoimmune disease; Parkinson's disease; silicosis;  
KW gastrointestinal disease; atherosclerosis; haemophilia; thrombocytopenia;  
KW immunosuppressive agent; adjuvant; enhance immune response;

KM higher affinity antibody induction;  
XX increased serum immunoglobulin concentration.

OS Homo sapiens.

XX US2002168711-A1.

XX 14-NOV-2002.

PF 17-JAN-2001; 2001US-00764868.

XX 31-JAN-2000; 2000US-0179065P.

PR 04-FEB-2000; 2000US-0180628P.

PR 28-JUN-2000; 2000US-0214886P.

PR 07-JUL-2000; 2000US-0216647P.

PR 07-JUL-2000; 2000US-0216880P.

PR 11-JUL-2000; 2000US-0217487P.

PR 11-JUL-2000; 2000US-0217496P.

PR 14-JUL-2000; 2000US-0218290P.

PR 26-JUL-2000; 2000US-0220963P.

PR 26-JUL-2000; 2000US-0220964P.

PR 14-AUG-2000; 2000US-0224518P.

PR 14-AUG-2000; 2000US-0224519P.

PR 14-AUG-2000; 2000US-0225267P.

PR 14-AUG-2000; 2000US-0225268P.

PR 14-AUG-2000; 2000US-0225270P.

PR 14-AUG-2000; 2000US-0225447P.

PR 14-AUG-2000; 2000US-0225757P.

PR 14-AUG-2000; 2000US-0225758P.

PR 22-AUG-2000; 2000US-0226868P.

PR 30-AUG-2000; 2000US-0228924P.

PR 01-SEP-2000; 2000US-0229287P.

PR 01-SEP-2000; 2000US-0229343P.

PR 01-SEP-2000; 2000US-0229344P.

PR 01-SEP-2000; 2000US-0229345P.

PR 05-SEP-2000; 2000US-0229509P.

PR 08-SEP-2000; 2000US-0229513P.

PR 21-SEP-2000; 2000US-0234223P.

PR 21-SEP-2000; 2000US-0234274P.

PR 25-SEP-2000; 2000US-0234997P.

PR 27-SEP-2000; 2000US-0235834P.

PR 29-SEP-2000; 2000US-0236337P.

PR 29-SEP-2000; 2000US-0236337P.

PR 29-SEP-2000; 2000US-0236368P.

PR 29-SEP-2000; 2000US-0236369P.

PR 29-SEP-2000; 2000US-0236370P.

PR 02-OCT-2000; 2000US-0236802P.

PR 02-OCT-2000; 2000US-0237037P.

PR 02-OCT-2000; 2000US-0237038P.

PR 02-OCT-2000; 2000US-0237039P.

PR 02-OCT-2000; 2000US-0237040P.

PR 13-OCT-2000; 2000US-0239935P.

PR 20-OCT-2000; 2000US-0240960P.

PR 20-OCT-2000; 2000US-0241785P.

PR 20-OCT-2000; 2000US-0241809P.

PR 01-NOV-2000; 2000US-0244617P.

PR 17-NOV-2000; 2000US-0249299P.

PR 08-DEC-2000; 2000US-0251856P.

PR 08-DEC-2000; 2000US-0251868P.

PR 08-DEC-2000; 2000US-0251869P.

XX (ROSE/) ROSEN C A.

PA (RUBEN/) RUBEN S M.

PA (BARA/) BARASH S C.

XX Rosen CA, Ruben SM, Barash SC;

XX WPI; 2003-719985/68.

DR N-PSDB; ADB93674.

XX New isolated polypeptide useful for diagnosing and treating

PT immunosuppressive conditions such as autoimmune disease and Parkinson's

PT disease.

XX Claim 11; SEQ ID NO 1154; 345pp; English.

XX The invention relates to an isolated polypeptide. The polypeptide is

CC useful for diagnosing a pathological condition or a susceptibility to a

CC pathological condition in a subject, by determining the presence or

CC amount of expression of the polypeptide in a biological sample and

CC diagnosing a pathological condition or a susceptibility to a pathological

CC condition based on the presence or amount of expression of the

CC polypeptide. The polypeptide is also useful for identifying a binding

CC partner to the polypeptide, which involves contacting the polypeptide

CC with a binding partner and determining whether the binding partner

CC affects an activity of the polypeptide. The polypeptide or the nucleic

CC acid encoding the polypeptide is useful for preventing, treating, or

CC ameliorating a medical condition, which involves administering the

CC polypeptide or the nucleic acid to a mammalian subject. The nucleic acid

CC is useful for diagnosing a pathological condition or a susceptibility to

CC a pathological condition in a subject, which involves determining the

CC presence or absence of a mutation in the nucleic acid, and diagnosing a

CC pathological condition or susceptibility to a pathological condition

CC based on the presence or absence of the mutation. The polypeptide, the

CC nucleic acid and an antibody to the polypeptide are useful for treating

CC autoimmune disease, Parkinson's disease, sickle cell, gastrointestinal

CC disease, atherosclerosis, haemophilia, thrombocytopenia. The polypeptide,

CC the nucleic acid and the antibody are useful as immunosuppressive agents,

CC as adjuvants to enhance immune responses, and as agents to induce higher

CC affinity antibodies and increase serum immunoglobulin concentrations. The

CC present sequence represents the amino acid sequence of a novel human

CC protein. Note: The sequence data for this patent did not form part of the

CC printed specification but was obtained in electronic format direct from

XX USPTO at seqdata.uspto.gov/sequence.html?DocId=20020168711.

XX SEQ Sequence 159 AA;

XX Query Match 25.7%; Score 733; DB 7; Length 159;

XX Best Local Similarity 90.6%; Pred. No. 1.9e-61;

XX Matches 144; Conservative 1; Mismatches 14; Indels 0; Gaps 0;

XX Db 2 ATRALSEELDNEDYSLNVRREASSEELKAYRRLCMYHPDKRDPDLKQAEKRLFTL 61

XX 1 ATRALSEELDNEDYSLNVRREASSEELKAYRRLCMYHPDKRDPDLKQAEKRLFTL 60

XX QY 62 VHQAYEVLSDPQTRAIYDIYGRKGLMEGMEVEVERRRTPAEIREEFERLQREERERLQ 121

XX Db 61 VHQAYEVLSDPQTRAIYDIYGRKGLMEGMEVEVERRRTPAEIREEFERLQREERERLQ 120

XX QY 122 RTNPKGTISVGVDAIDLFDRYDEYEDVSGSFPQIETN 160

XX Db 121 RTNPKGTISVGVDAIDLFDRYDXLXLMCPAVXFPYIEIN 159

XX RESULT 14

XX ABG14695

XX ID ABG14695 standard; protein; 147 AA.

XX XX

XX AC ABG14695;

XX XX

XX XX 18-FEB-2002 (first entry)

XX XX

XX DE Novel human diagnostic protein #14686.

XX XX

XX KM Human: chromosome mapping; gene mapping; gene therapy; forensic;

XX KM food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX XX

XX XX WO200175067-A2.

XX XX

XX PD 11-OCT-2001.

XX XX

XX PF 30-MAR-2001; 2001WO-US008631.

XX XX

PR 31-MAR-2000; 2000US-00540217.  
PR 23-AUG-2000; 2000US-00649167.  
XX (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI: 2001-639362/73.  
DR N-PSDB; AAS78882.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.  
XX  
PS Claim 20; SEQ ID NO 45054; 103pp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and polypeptide (II)  
CC sequences. (I) is useful as hybridization probes, polymerase chain  
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
CC and in recombinant production of (II). The polynucleotides are also used  
CC in diagnostics as expressed sequence tags for identifying expressed  
CC genes. (I) is useful in gene therapy techniques to restore normal  
CC activity of (II) or to treat disease states involving (II). (II) is  
CC useful for generating antibodies against it, detecting or quantitating a  
CC polypeptide in tissue, as molecular weight markers and as a food  
CC supplement. (II) and its binding partners are useful in medical imaging  
CC of sites expressing (II). (I) and (II) are useful for treating disorders  
CC involving aberrant protein expression or biological activity. The  
CC polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic  
CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic  
CC patent did not appear in the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 147 AA;  
Query Match 23.7%; Score 676.5; DB 4; Length 147;  
Best Local Similarity 84.0%; Pred. No. 4.1e-56;  
Matches 137; Conservative 0; Mismatches 1; Indels 25; Gaps 1;  
QY 4 ALSEEBELDNEDYSLNVRREASSELKAYRRLCMLYHPDKRDPKLSQAERLFN 63  
DB 10 ALSEEBELDNEDYSLNVRREASSELKAYRRLCMLYHPDKRDPKLSQAERLFN 69  
QY 64 QAYEVLSDPQTRAIYDIYGRKGLMEGMEVVERRRTPAEIRREFFERLQRRERRL 123  
DB 70 QAY-----EVERRRTPAEIRREFFERLQRRERRLQORT 104  
QY 124 NPKGTISGVADTLPFRYDEYEDVSGSFPQIENKMHISQ 166  
DB 105 NPKGTISGVADTLPFRYDEYEDVSGSFPQIENKMHISQ 147  
RESULT 15  
AAG01815  
ID AAG01815 standard; protein; 119 AA.  
XX  
AC AAG01815;  
XX  
DT 06-OCT-2000 (first entry)  
XX  
DE Human secreted protein, SEQ ID NO: 5896.  
XX  
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
XX gene therapy; chromosome mapping.  
XX  
OS Homo sapiens.  
XX

PN EP1033401-A2.  
PD 06-SEP-2000.  
XX  
PD 21-FEB-2000; 2000EP-00200610.  
XX  
PF 26-FEB-1999; 99US-0122487P.  
XX  
PR (BEST) GENSET.  
XX  
PI Dumas Milne Edwards J, Duclert A, Giordano J;  
XX  
DR WPI: 2000-500381/45.  
DR N-PSDB; AAC01821.  
XX  
XX  
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for  
PT diagnostic, forensic, gene therapy and chromosome mapping procedures.  
XX  
PS Claim 13; SEQ ID NO 5896; 71pp + Sequence Listing; English.  
XX  
CC The present sequence is a polypeptide encoded by one of a large number of  
CC 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs were  
CC prepared from total human RNAs or polyA+ RNAs derived from 30 different  
CC tissues. EST sequences usually correspond mainly to the 3' untranslated  
CC region (UTR) of the mRNA because they are often obtained from oligo-dT  
CC primed cDNA libraries. Such ESTs are not well suited for isolating cDNA  
CC sequences derived from the 5' ends of mRNAs and even in those cases where  
CC longer cDNA sequences have been obtained, the full 5' UTR is rarely  
CC included. 5' ESTs are derived from mRNAs with intact 5' ends and can  
CC therefore be used to obtain full length cDNAs and genomic DNAs. 5' ESTs  
CC are also used in diagnostic, forensic, gene therapy and chromosome  
CC mapping procedures. They are used to obtain upstream regulatory sequences  
CC and to design expression and secretion vectors  
XX  
SQ Sequence 119 AA;  
Query Match 21.5%; Score 613; DB 3; Length 119;  
Best Local Similarity 100.0%; Pred. No. 3.4e-50;  
Matches 119; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MATALSEEBELDNEDYSLNVRREASSELKAYRRLCMLYHPDKRDPKLSQAERLFN 60  
DB 1 MATALSEEBELDNEDYSLNVRREASSELKAYRRLCMLYHPDKRDPKLSQAERLFN 60  
QY 61 LVHQAAYEVLSDPQTRAIYDIYGRKGLMEGMEVVERRRTPAEIRREFFERLQRRERRL 119  
DB 61 LVHQAAYEVLSDPQTRAIYDIYGRKGLMEGMEVVERRRTPAEIRREFFERLQRRERRL 119  
Search completed: July 13, 2004, 08:40:48  
Job time : 86 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 13, 2004, 08:39:16 ; Search time 24 Seconds  
(without alignments)  
1202.455 Million cell updates/sec

Title: US-09-787-678A-6

Perfect score: 2857  
Sequence: 1 MATALSEEDLNEDYSLN.....VLDEALRIPKQSHRIDTG 559

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA: \*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep: \*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep: \*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep: \*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep: \*  
5: /cgn2\_6/ptodata/2/1aa/PTCUS.COMB.pep: \*  
6: /cgn2\_6/ptodata/2/1aa/backfile1.pep: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	223	7.8	330	2	US-08-868-288A-3
2	223	7.8	330	3	US-09-235-373-3
3	223	7.8	330	3	US-09-388-993-3
4	195	6.8	392	4	US-09-198-452A-48
5	194	6.8	340	2	US-08-974-546-5
6	192	6.7	337	4	US-09-665-479A-8
7	192	6.7	348	4	US-08-974-546-1
8	189	6.6	87	4	US-09-882-835-4
9	187	6.5	277	2	US-08-868-288A-7
10	187	6.5	277	3	US-09-235-373-7
11	187	6.5	277	3	US-09-388-993-7
12	187	6.5	351	2	US-08-868-288A-6
13	187	6.5	351	3	US-09-388-993-6
14	187	6.5	351	3	US-09-388-993-6
15	186	6.5	153	4	US-09-621-976-5129
16	185	6.5	381	4	US-09-253-991A-27174
17	181	6.3	380	4	US-09-488-039A-9011
18	180.5	6.3	419	2	US-08-686-417-3
19	177.5	6.2	399	4	US-09-553-498-2
20	177.5	6.2	399	4	US-09-618-869-2
21	176.5	6.2	380	4	US-09-543-681A-5912
22	175.5	6.1	131	4	US-09-553-498-4
23	175.5	6.1	131	4	US-09-618-869-4
24	171.5	6.0	375	4	US-09-328-352-4984
25	169	5.9	352	2	US-08-472-534-6
26	166	5.8	127	4	US-09-370-838-199
27	166	5.8	397	2	US-08-868-288A-5

28	166	5.8	397	3	US-09-235-373-5	Sequence 5, Appli
29	166	5.8	397	3	US-09-388-993-5	Sequence 5, Appli
30	164.5	5.8	358	2	US-08-868-288A-1	Sequence 1, Appli
31	164.5	5.8	358	3	US-09-235-373-1	Sequence 1, Appli
32	164.5	5.8	358	3	US-09-388-993-1	Sequence 1, Appli
33	161.5	5.7	385	4	US-09-134-001C-3688	Sequence 1, Appli
34	160.5	5.6	332	4	US-09-882-835-2	Sequence 2, Appli
35	160.5	5.6	407	4	US-09-540-236-2899	Sequence 2899, Ap
36	158.5	5.5	320	4	US-09-328-352-7937	Sequence 7937, Ap
37	156	5.5	438	2	US-08-897-340-34	Sequence 34, Appli
38	156	5.5	438	3	US-09-252-329-34	Sequence 34, Appli
39	156	5.5	484	2	US-08-879-260-4	Sequence 4, Appli
40	156	5.5	484	3	US-09-231-529-4	Sequence 4, Appli
41	156	5.5	484	3	US-08-977-816-4	Sequence 4, Appli
42	153	5.4	223	4	US-09-658-544-4	Sequence 4, Appli
43	151	5.3	504	3	US-09-231-529-6	Sequence 6, Appli
44	151	5.3	504	3	US-08-977-816-6	Sequence 6, Appli
45	150.5	5.3	320	4	US-09-543-681A-7884	Sequence 7884, Ap

## ALIGNMENTS

RESULT 1  
US-08-868-288A-3  
Sequence 3, Application US/0868288A  
Patent No. 5922567  
GENERAL INFORMATION:  
APPLICANT: Au-Young, Janice  
APPLICANT: Lal, Preeti  
APPLICANT: Bandman, Olga  
TITLE OF INVENTION: TWO NEW HUMAN DNA-LIKE PROTEINS  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESSES:  
ADDRESS: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/868,288A  
FILING DATE: June 3, 1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0309 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 330 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: HNT2RAT01  
CLONE: 260873  
US-08-868-288A-3

Query Match 7.8%; Score 223; DB 2; Length 330;  
Best Local Similarity 32.1%; Pred. No. 1.8e-14;  
Matches 67; Conservative 22; Mismatches 50; Indels 70; Gaps 7;







APPLICANT: Bandaru, Rajasekhar

```
IOPOLOGI: linear
IMMEDIATE SOURCE:
```

```

; IMMEDIATE SOURCE:
;
; LIBRARY: GenBank

```







***This Page Blank (uspiu)***

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 13, 2004, 08:41:11 ; Search time 352 Seconds  
(without alignments)  
495.347 Million cell updates/sec

Title: US-09-787-678A-6

Perfect score: 2857  
Sequence: 1 MATLSEEDLNEDYSLN.....VLDSSELRIRKSHRIDTDG 559

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1279676 seqs, 311918243 residues

Total number of hits satisfying chosen parameters: 1279676

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Published Applications AA:\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*
- 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep:\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep:\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep:\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep:\*
- 6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep:\*
- 7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep:\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep:\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep:\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep:\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep:\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep:\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep:\*
- 15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep:\*
- 16: /cgn2\_6/ptodata/1/pubpaa/US10C\_NEW\_PUB.pep:\*
- 17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep:\*
- 18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2857	100.0	559	US-10-000-897-70	Sequence 70, App1
2	2857	100.0	559	US-10-408-765A-1911	Sequence 911, App
3	2848	99.7	559	US-10-408-765A-941	Sequence 941, App1
4	1723	60.3	336	US-10-000-897-69	Sequence 69, App1
5	750	26.3	159	US-09-764-868-780	Sequence 780, App
6	733	25.7	159	US-09-764-868-1154	Sequence 1154, App
7	661	23.1	569	US-10-437-963-131723	Sequence 131723, App
8	456.5	16.0	118	US-09-927-738-17	Sequence 17, App1
9	451	15.8	103	US-10-000-897-71	Sequence 71, App1
10	364	12.7	338	US-10-424-599-225707	Sequence 225707, App
11	314	11.0	234	US-10-424-599-184546	Sequence 184546, App
12	300.5	10.5	322	US-10-425-114-71459	Sequence 71459, App
13	239.5	8.4	369	US-10-369-493-2975	Sequence 2975, App
14	223	7.8	242	US-10-264-049-2674	Sequence 2674, App
15	222.5	7.8	376	US-10-369-493-47	Sequence 47, App1

16	220.5	7.7	342	US-10-437-963-126854	Sequence 126854, App
17	217	7.6	158	US-10-425-114-48912	Sequence 48912, App
18	213	7.5	116	US-10-424-599-233524	Sequence 233524, App
19	210.5	7.4	340	US-10-424-599-278129	Sequence 278129, App
20	210.5	7.4	362	US-10-437-963-190831	Sequence 190831, App
21	210	7.4	146	US-10-424-599-191916	Sequence 191916, App
22	208	7.2	153	US-10-424-599-257493	Sequence 257493, App
23	206	7.2	404	US-10-425-114-59964	Sequence 59964, App
24	205.5	7.2	356	US-10-437-963-126176	Sequence 126176, App
25	204	7.1	192	US-10-424-599-237535	Sequence 237535, App
26	203	7.1	545	US-10-437-963-114080	Sequence 114080, App
27	202	7.1	217	US-10-424-599-257471	Sequence 257471, App
28	198	6.9	411	US-10-425-114-46080	Sequence 46080, App
29	197	6.9	289	US-10-424-599-192793	Sequence 192793, App
30	196	6.9	352	US-10-425-114-37469	Sequence 37469, App
31	195.5	6.8	276	US-10-424-599-191915	Sequence 191915, App
32	195	6.8	392	US-10-289-762-48	Sequence 48, App1
33	194	6.8	340	US-10-116-275-134	Sequence 134, App
34	194	6.8	340	US-10-116-275-135	Sequence 135, App
35	194	6.8	375	US-10-369-493-18988	Sequence 18988, App
36	194	6.8	433	US-10-425-114-60908	Sequence 60908, App
37	193	6.8	383	US-10-437-963-108162	Sequence 108162, App
38	193	6.8	440	US-10-424-599-253715	Sequence 253715, App
39	193	6.8	478	US-10-425-114-65215	Sequence 65215, App
40	192	6.7	433	US-10-425-114-63266	Sequence 63266, App
41	192	6.7	435	US-10-425-114-59462	Sequence 59462, App
42	191	6.7	288	US-10-369-493-8874	Sequence 8874, App
43	191	6.7	334	US-10-161-051-134	Sequence 134, App
44	191	6.7	435	US-10-425-114-65366	Sequence 65366, App
45	190	6.7	365	US-10-369-493-13503	Sequence 13503, App

ALIGNMENTS

RESULT 1	US-10-000-897-70	Application US/10000897
Sequence 70, App1	US-10-000-897-70	Application US/10000897
Publicatation No. US20030165852A1		
GENERAL INFORMATION:		
APPLICANT: Schellert, P.		
TITLE OF INVENTION: METHODS AND REAGENTS FOR IDENTIFYING RARE FETAL CELLS IN THE		
FILE REFERENCE: 11012-004-999		
CURRENT APPLICATION NUMBER: US/10/000,897		
CURRENT FILING DATE: 2001-11-15		
PRIOR APPLICATION NUMBER: 60/248,882		
PRIOR FILING DATE: 2000-11-15		
NUMBER OF SEQ ID NOS: 78		
SOFTWARE: PatentIn version 3.0		
SEQ ID NO 70		
LENGTH: 559		
TYPE: PRT		
ORGANISM: Homo sapiens		
US-10-000-897-70		
Query Match	100.0%;	Score 2857; DB 14; length 559;
Best Local Similarity	100.0%;	Pred. No. 3.9e-263;
Matches 559; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
Qy	1	MATLSEEDLNEDYSLNVRREASSEELKAAYRRCMLYHPDKARDPELKSQAERLFN 60
Db	1	MATLSEEDLNEDYSLNVRREASSEELKAAYRRCMLYHPDKARDPELKSQAERLFN 60
Qy	61	LHVQAYVLSDPOTRAIYDYGKRGLEMEGMEVVERRTPAJIRBEFERLORREBERRLQ 120
Db	61	LHVQAYVLSDPOTRAIYDYGKRGLEMEGMEVVERRTPAJIRBEFERLORREBERRLQ 120
Qy	121	QRTNPKCTISGVADATLDFRDYDEYEDVSGSFPQIENKMIQSISIAPLATDTAIL 180
Db	121	QRTNPKCTISGVADATLDFRDYDEYEDVSGSFPQIENKMIQSISIAPLATDTAIL 180
Qy	181	SGSLSTONGNGGGSINFALRKVTSAKGMGELFGAGLQPLFGKLFRULTRCFVTYN 240

```
Db 181 SGLSTQNGNGGSSINPALRRVTSAGKMGELFAGADLOGPLFGKLPRLTFRCEVTNN 240
Qy 241 CALQSSRGIRPGLTTVLARNLDKNTVGYLQWRMGISAMNTSIYRDTKTSHTVALQLG 300
Db 241 CALQSSRGIRPGLTTVLARNLDKNTVGYLQWRMGISAMNTSIYRDTKTSHTVALQLG 300
Qy 301 IPHSPALISYQHKFODDDQTRVKGSLKAGFGTVVEYGARRKISRHSVLGAASVGVPOG 360
Db 301 IPHSPALISYQHKFODDDQTRVKGSLKAGFGTVVEYGARRKISRHSVLGAASVGVPOG 360
Qy 361 VSLKVLNRAASQTYFFPHLTDQLPSAMFYATVGPLVYFAMHRLIIKPYLRAQEKEL 420
Db 361 VSLKVLNRAASQTYFFPHLTDQLPSAMFYATVGPLVYFAMHRLIIKPYLRAQEKEL 420
Qy 421 EKQRESAATDVLOKQOEAESAVALMOESVRIIEAESRMGLIIVNAYGKFPVNDKSRKS 480
Db 421 EKQRESAATDVLOKQOEAESAVALMOESVRIIEAESRMGLIIVNAYGKFPVNDKSRKS 480
Qy 481 EKQKVIDVTVPLQCLVKOSKLLITBASKAGLPGFYPCVGEENKLVLYQFRGVLHQVW 540
Db 481 EKQKVIDVTVPLQCLVKOSKLLITBASKAGLPGFYPCVGEENKLVLYQFRGVLHQVW 540
Qy 541 LDSEALRIPKQSHRIDTDG 559
Db 541 LDSEALRIPKQSHRIDTDG 559
```

## RESULT 2

```
US-10-408-765A-1911
; Sequence 1911, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Bojin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1911
; LENGTH: 559
; TYPE: PRF
; ORGANISM: Homo sapiens
US-10-408-765A-1911
```

```
Query Match 100.0%; Score 2857; DB 16; Length 559;
Best Local Similarity 100.0%; Pred. No. 3.9e-263;
Matches 559; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1 MATALSEEBELNDYISLNVREASSELKAAVRLCMLYHPDKRDELSQAERLPN 60
Db 1 MATALSEEBELNDYISLNVREASSELKAAVRLCMLYHPDKRDELSQAERLPN 60
Qy 61 LVHQAIEVLSDPQTRAIYDIYGRGLMEGMEVEVERRRTPAIRREPERLQREBERRLQ 120
Db 61 LVHQAIEVLSDPQTRAIYDIYGRGLMEGMEVEVERRRTPAIRREPERLQREBERRLQ 120
Qy 121 QRTNPKGITSVGVADATDLFDRYDEEYEDVSGSPQIEINXKHISQSIAPLTATDTALL 180
Db 121 QRTNPKGITSVGVADATDLFDRYDEEYEDVSGSPQIEINXKHISQSIAPLTATDTALL 180
Qy 181 SGLSTQNGNGGSSINPALRRVTSAGKMGELFAGADLOGPLFGKLPRLTFRCEVTNN 240
Db 181 SGLSTQNGNGGSSINPALRRVTSAGKMGELFAGADLOGPLFGKLPRLTFRCEVTNN 240
```

```
Qy 241 CALQSSRGIRPGLTTVLARNLDKNTVGYLQWRMGISAMNTSIYRDTKTSHTVALQLG 300
Db 241 CALQSSRGIRPGLTTVLARNLDKNTVGYLQWRMGISAMNTSIYRDTKTSHTVALQLG 300
Qy 301 IPHSPALISYQHKFODDDQTRVKGSLKAGFGTVVEYGARRKISRHSVLGAASVGVPOG 360
Db 301 IPHSPALISYQHKFODDDQTRVKGSLKAGFGTVVEYGARRKISRHSVLGAASVGVPOG 360
Qy 361 VSLKVLNRAASQTYFFPHLTDQLPSAMFYATVGPLVYFAMHRLIIKPYLRAQEKEL 420
Db 361 VSLKVLNRAASQTYFFPHLTDQLPSAMFYATVGPLVYFAMHRLIIKPYLRAQEKEL 420
Qy 421 EKQRESAATDVLOKQOEAESAVALMOESVRIIEAESRMGLIIVNAYGKFPVNDKSRKS 480
Db 421 EKQRESAATDVLOKQOEAESAVALMOESVRIIEAESRMGLIIVNAYGKFPVNDKSRKS 480
Qy 481 EKQKVIDVTVPLQCLVKOSKLLITBASKAGLPGFYPCVGEENKLVLYQFRGVLHQVW 540
Db 481 EKQKVIDVTVPLQCLVKOSKLLITBASKAGLPGFYPCVGEENKLVLYQFRGVLHQVW 540
Qy 541 LDSEALRIPKQSHRIDTDG 559
Db 541 LDSEALRIPKQSHRIDTDG 559
```

## RESULT 3

```
US-10-408-765A-941
; Sequence 941, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Bojin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 941
; LENGTH: 559
; TYPE: PRF
; ORGANISM: Homo sapiens
US-10-408-765A-941
```

```
Query Match 99.7%; Score 2848; DB 16; Length 559;
Best Local Similarity 99.6%; Pred. No. 2.8e-262;
Matches 557; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```

```
Qy 1 MATALSEEBELNDYISLNVREASSELKAAVRLCMLYHPDKRDELSQAERLPN 60
Db 1 MATALSEEBELNDYISLNVREASSELKAAVRLCMLYHPDKRDELSQAERLPN 60
Qy 61 LVHQAIEVLSDPQTRAIYDIYGRGLMEGMEVEVERRRTPAIRREPERLQREBERRLQ 120
Db 61 LVHQAIEVLSDPQTRAIYDIYGRGLMEGMEVEVERRRTPAIRREPERLQREBERRLQ 120
Qy 121 QRTNPKGITSVGVADATDLFDRYDEEYEDVSGSPQIEINXKHISQSIAPLTATDTALL 180
Db 121 QRTNPKGITSVGVADATDLFDRYDEEYEDVSGSPQIEINXKHISQSIAPLTATDTALL 180
Qy 181 SGLSTQNGNGGSSINPALRRVTSAGKMGELFAGADLOGPLFGKLPRLTFRCEVTNN 240
Db 181 SGLSTQNGNGGSSINPALRRVTSAGKMGELFAGADLOGPLFGKLPRLTFRCEVTNN 240
Qy 241 CALQSSRGIRPGLTTVLARNLDKNTVGYLQWRMGISAMNTSIYRDTKTSHTVALQLG 300
Db 241 CALQSSRGIRPGLTTVLARNLDKNTVGYLQWRMGISAMNTSIYRDTKTSHTVALQLG 300
```



```
Qy 301 IPHSFALISYOHKFQDDQTRVKSLKAGFGTVVEGAERKISRHSVLGAASVGVPOG 360
Db 301 IPHSFALISYOHKFQDDQTRVKSLKAGFGTVVEGAERKISRHSVLGAASVGVPOG 360
Qy 361 VSLKVKLNRAASQTYFFPIHLTDQLPSAMFYATGVVVYPAMRLIIKPYLRAQKEKEL 420
Db 361 VSLKVKLNRAASQTYFFPIHLTDQLPSAMFYATGVVVYPAMRLIIKPYLRAQKEKEL 420
Qy 421 EKORSEATDVLYQKKOAEASAVRLMOGSVRRRIIEAESRMGLIIVNAMYGFVNDKSRKS 480
Db 421 EKORSEATDVLYQKKOAEASAVRLMOGSVRRRIIEAESRMGLIIVNAMYGFVNDKSRKS 480
Qy 481 EKXVIVDVTVPLQCLVNDKSKILLTEASKAGLPGFYDPCVGEERKLYLXQPRGLHGVNV 540
Db 481 EKXVIVDVTVPLQCLVNDKSKILLTEASKAGLPGFYDPCVGEERKLYLXQPRGLHGVNV 540
Qy 541 LDSEALRIPKQSHRIDTDG 559
Db 541 LDSEALRIPKQSHRIDTDG 559
```

```
RESULT 4
US-10-000-897-69
; Sequence 69, Application US/10000897
; Publication No. US20030165852A1
; GENERAL INFORMATION:
; APPLICANT: Schueler, P.
; TITLE OF INVENTION: METHODS AND REAGENTS FOR IDENTIFYING RARE FETAL CELLS IN THE
; FILE REFERENCE: 11012-004-999
; CURRENT APPLICATION NUMBER: US/10/000,897
; PRIOR FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: 60/248,862
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 69
; LENGTH: 336
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-000-897-69
```

```
Query Match 60.3%; Score 1723; DB 14; Length 336;
Best Local Similarity 99.1%; Pred. No. 2,6e-155;
Matches 333; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
```

```
Qy 1 MATALSEELDNEDYSLNVRREASSEBELKAAYRLCMLYHPDKHRDPELKSQAERLFLN 60
Db 1 MATALSEELDNEDYSLNVRREASSEBELKAAYRLCMLYHPDKHRDPELKSQAERLFLN 60
Qy 61 LVHQAAYEVLSDPQTRAIYDIYGRKGLMEGWEVVERRTTAEIRREPERLQREBERRLQ 120
Db 61 LVHQAAYEVLSDPQTRAIYDIYGRKGLMEGWEVVERRTTAEIRREPERLQREBERRLQ 120
Qy 121 ORTNPKGTISVGVDATDLFDRYDEYEDVSGSSPQLEINKMHSQSIETAPLTATDAIL 180
Db 121 ORTNPKGTISVGVDATDLFDRYDEYEDVSGSSPQLEINKMHSQSIETAPLTATDAIL 180
Qy 181 SGLSLTONGNGGSGSINFALRRVTSKAGMGELEFGAGLOGPLFGKLFRMLTPRCFVTYN 240
Db 181 SGLSLTONGNGGSGSINFALRRVTSKAGMGELEFGAGLOGPLFGKLFRMLTPRCFVTYN 240
Qy 241 CALQFSRGRIRPGITVTLARNLDKNTVGYLQWRMGIOSAMNTSIVRDTKTSHTVALQLG 300
Db 241 CALQFSRGRIRPGITVTLARNLDKNTVGYLQWRMGIOSAMNTSIVRDTKTSHTVALQLG 300
Qy 301 IPHSFALISYOHKFQDDQTRVKSLKAGFGTVVE 336
Db 301 IPHSFALISYOHKFQDDQTRVKSLKAGFGTVVE 336
```

RESULT 5

```
US-09-764-868-780
; Sequence 780, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT232
; CURRENT APPLICATION NUMBER: US/09/764,868
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION DATA REMOVED - REFER TO PALM OR FILE WRAPPER
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 780
; LENGTH: 159
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (107)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (143)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (144)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (152)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (155)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-868-780
```

```
Query Match 26.3%; Score 750; DB 9; Length 159;
Best Local Similarity 92.5%; Pred. No. 5.1e-63;
Matches 147; Conservative 1; Mismatches 11; Indels 0; Gaps 0;
```

```
Qy 2 ATALSEELDNEDYSLNVRREASSEBELKAAYRLCMLYHPDKHRDPELKSQAERLFLN 61
Db 1 ATALSEELDNEDYSLNVRREASSEBELKAAYRLCMLYHPDKHRDPELKSQAERLFLN 60
Qy 62 VHQAYEVLSDPQTRAIYDIYGRKGLMEGWEVVERRTTAEIRREPERLQREBERRLQ 121
Db 62 VHQAYEVLSDPQTRAIYDIYGRKGLMEGWEVVERRTTAEIRREPERLQREBERRLQ 120
Qy 122 RTNPKGTISVGVDATDLFDRYDEYEDVSGSSPQLEIN 160
Db 122 RTNPKGTISVGVDATDLFDRYDXKLKMCRAVXPEPIEIN 159
```

```
RESULT 6
US-09-764-868-1154
; Sequence 1154, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT232
; CURRENT APPLICATION NUMBER: US/09/764,868
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION DATA REMOVED - REFER TO PALM OR FILE WRAPPER
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1154
; LENGTH: 159
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (31)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
```

LOCATION: (58)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (105)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (107)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (143)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (144)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (152)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (155)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-764-868-1154

Query Match 25.7%; Score 733; DB 9; Length 159;  
Best Local Similarity 90.6%; Pred. No. 2,2e-61;  
Matches 144; Conservative 1; Mismatches 14; Indels 0; Gaps 0;

QY 2 ATALSEELDNEDYSLNVRREASSEELKAAVRLCMLYHPDKRDPKLSQARLFLNL 61  
DB 1 ATALSEELDNEDYSLNVRREASSEELKAAVRLCMLYHPDKRDPKLSQARLFLNL 60  
QY 62 VHQAYEVLSDPQTRAIYDIYGRKGLMEGMEVEVERRRTPAIREPERLQREBERRLQ 121  
DB 61 VHQAYEVLSDPQTRAIYDIYGRKGLMEGMEVEVERRRTPAIREPERLQREBERRLQ 120  
QY 122 RINPKGTISGVNATDLPFRYDEYEDVSSSPQIEIN 160  
DB 121 RINPKGTISGVNATDLPFRYDXKLMCPAVXFPXIEIN 159

## RESULT 7

US-10-437-963-131723  
Sequence 131723, Application US/10437963  
Publication No. US2004013343A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa, Thomas J.  
APPLICANT: Kovalic, David K.  
APPLICANT: Zhou, Yihua  
APPLICANT: Cao, Yongwei  
APPLICANT: Wu, Wei  
APPLICANT: Boukharov, Andrey A.  
APPLICANT: Barbazuk, Brad  
APPLICANT: Li, Ping  
TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(53221) B  
CURRENT FILING DATE: 2003-05-14  
NUMBER OF SEQ ID NOS: 204966  
SEQ ID NO 131723  
LENGTH: 569  
TYPE: PRT  
ORGANISM: Oryza sativa  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT4530\_33762C.1.pep  
US-10-437-963-131723

Query Match 23.1%; Score 661; DB 16; Length 569;  
Best Local Similarity 30.2%; Pred. No. 1,2e-53;  
Matches 178; Conservative 103; Mismatches 242; Indels 66; Gaps 14;

QY 8 EELDNDYSLNVRREASSEELKAAVRLCMLYHPDKRDPKLSQARLFLNLVHQAYE 67  
DB 6 EELDNDYSLNVRREASSEELKAAVRLCMLYHPDKRDPKLSQARLFLNLVHQAYE 65

QY 68 VLSDPQTRAIYDIYGRKGLMEGMEVEVERRRTPAIREPERLQREBERRLQORTNPKG 127  
DB 66 VLSDPQTRAIYDIYGRKGLMEGMEVEVERRRTPAIREPERLQREBERRLQORTNPKG 124  
QY 128 TISGVNATDLPFRYDEYEDVSSSPQIEINPKHISQISAPPLATDAILSGSLSTQ 187  
DB 125 TISGVNATDLPFRYDEYEDVSSSPQIEINPKHISQISAPPLATDAILSGSLSTQ 170  
QY 188 NNGGSGINPALARITSAKMGELFPGADLOGPFGKLFNLPVPCVTTNCLQPS 247  
DB 171 NNGGSGINPALARITSAKMGELFPGADLOGPFGKLFNLPVPCVTTNCLQPS 228  
QY 248 RGRPGLATVILARNLDKNTVGYLQWRMGISAMNTSIYVDTISFTVALQGIPIHSPAL 307  
DB 223 RGRPGLATVILARNLDKNTVGYLQWRMGISAMNTSIYVDTISFTVALQGIPIHSPAL 286  
QY 308 ISYQHKFODDQTRVYSKAGPFGTVYGAERKISRHSVLGAASVGVV----- 358  
DB 287 ISYQHKFODDQTRVYSKAGPFGTVYGAERKISRHSVLGAASVGVV----- 344  
QY 359 -----QVSLKVLNRAQTFFPIHLT---DQLPSAMFYATVCP 396  
DB 345 VINKGMIMKESIVIGVEGVSMPFELHRAQKLVIPVLTSTDPNALLATSVFAL---P 401  
QY 397 LVVYFAMRLIKPY-LRAQKELEKORESATDVLOKORAESAVRLMOESVRITEA 455  
DB 402 STLYPLQVYFVPCVCKREKKELEK-ESLSQLETRAKKAAQKILLEVSRNKKR 460  
QY 456 EESRMGLIIVNMYG-KFVNDKSRKSE-----KVKVIDVTVPLQCLYKDS-KLILTEA- 506  
DB 461 QEDGLVITVKALYGNRKVKESSENELNDVSAQVLDVTIPLNPLVSEAGQLKLHGI 520  
QY 507 SKAGLPGFVPCVGEERKULKLYQFRGVLYHQVMDSEALRIPKSHRI 555  
DB 521 SKAGLPGFVPCVGEERKULKLYQFRGVLYHQVMDSEALRIPKSHRI 569

## RESULT 8

US-09-927-738-17  
Sequence 17, Application US/09927738  
Patent No. US20020076799A1  
GENERAL INFORMATION:  
APPLICANT: Wang, Tongwen  
TITLE OF INVENTION: Compositions and Methods of modulating TGF-B signaling  
FILE REFERENCE: 17633/1082  
CURRENT APPLICATION NUMBER: US/09/927,738  
PRIOR FILING DATE: 2001-08-10  
PRIOR APPLICATION NUMBER: US 60/119786  
PRIOR FILING DATE: 1999-02-11  
NUMBER OF SEQ ID NOS: 28  
SOFTWARE: Patent version 3.1  
SEQ ID NO 17  
LENGTH: 118  
TYPE: PRT  
ORGANISM: Unknown  
FEATURE:  
OTHER INFORMATION: Clone S3 + 125 Protein Sequence  
US-09-927-738-17

Query Match 16.0%; Score 456.5; DB 9; Length 118;  
Best Local Similarity 83.5%; Pred. No. 3,2e-35;  
Matches 91; Conservative 6; Mismatches 11; Indels 1; Gaps 1;

QY 451 RIIAEESRMGLIIVNMYGKFNKSRKSEKVIIVTVPLQCLVYKSKLITREASRAG 510  
DB 11 RIIAEESRMGLIIVNMYGKFNKSRKSEKVIIVTVPLQCLVYKSKLITREASRAG 69  
QY 511 LFGFYDPCVGEERKULKLYQFRGVLYHQVMDSEALRIPKSHRIDTDG 559  
DB 70 LFGFYDPCVGEERKULKLYQFRGVLYHQVMDSEALRIPKSHRIDTDG 118

```
RESULT 9
US-10-000-897-71
; Sequence 71, Application US/10000897
; Publication No. US20030165852A1
; GENERAL INFORMATION:
; APPLICANT: Schueler, P.
; TITLE OF INVENTION: METHODS AND REAGENTS FOR IDENTIFYING RARE FETAL CELLS IN THE
; FILE REFERENCE: 11012-004-999
; CURRENT APPLICATION NUMBER: US/10/000,897
; CURRENT FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: 60/248,882
; PRIOR FILING DATE: 2000-11-15
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 71
; LENGTH: 103
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-000-897-71

Query Match          15.8%; Score 451; DB 14; Length 103;
Best Local Similarity 89.3%; Pred. No. 8,66-35;
Matches 92; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 79 DIYKRGLEMEGWEVERRRTPAIREPERLQREERERLQORTNPKGTISGVGDATDL 138
D 1 DVSXKKKKKKHKKVRRRTPAIREPERLQREERERLQORTNPKGTISGVGDATDL 60
QY 139 FDRYDEEYDVSSGSPFOIEINKMHISOSTEAPLTADTDL 181
D 61 FDRYDEEYDVSSGSPFOIEINKMHISOSTEAPLTADTDL 103

RESULT 10
US-10-424-599-225707
; Sequence 225707, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 225707
; LENGTH: 338
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(338)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_45845C.1.pep
US-10-424-599-225707

Query Match          12.7%; Score 364; DB 12; Length 338;
Best Local Similarity 27.7%; Pred. No. 1,1e-25;
Matches 106; Conservative 54; Mismatches 146; Indels 76; Gaps 11;

QY 8 EEIDNEHYVSLIWRRASSSEELKAAYRLCMLYHPKMDPELKSQAEFLFNLVHQAYE 67
D 6 EAQONRFLVALLNISPASDEIRRAYQMAQYHPKQYAPPHKDIATENFORICAYE 65
QY 68 VLSDPQTRAIYDIYKRGLEMEGWEVERRRTPAIREPERLQREERERLQORTNPKG 127
D 68 VLSDPQTRAIYDIYKRGLEMEGWEVERRRTPAIREPERLQREERERLQORTNPKG 127
```

```
US-10-424-599-184546
; Sequence 184546, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 184546
; LENGTH: 234
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_137661C.1.pep
US-10-424-599-184546

Query Match          11.0%; Score 314; DB 12; Length 234;
Best Local Similarity 39.0%; Pred. No. 3,7e-21;
Matches 90; Conservative 33; Mismatches 96; Indels 12; Gaps 7;

QY 335 VEYGAERKISRHSVLGAASVGVPGVSLKVKLNRASTQYFPPIHLTDOLPSAMFYATV 394
D 6 VEVGGGRKLSKFSFSSVWMLYIGI-QGISWFEIYRGQKLIIPILTRHILNPFVATGAFV 64
QY 395 GELVYFAMHRLIKR-YLRQKEKELEKOREAANDVLOKQGEASAVRLMESVRRRI 453
D 65 VPASLYFVLRKQLFKIKRYLRNRQKALR-EKERTSAQVYEAARYAKAQNLOQNVANRRK 123
QY 454 EAEESRMGLITIVAMYG--KFVNDKSRKSE-----KRVYIDVVPLOCTLVKDS-KLILTE 505
D 124 NKQLETFGGLVIMALYGNQRIINNLKSSSETSELTSEVYIDVYIPLNPLVNDGQKLTHE 183
QY 506 -ASKAGLPGFYDPCVGEKRLKVLQYFRGLVQVMVLDSBALRIPKQSHRI 555
D 184 GVKSGSIGMFCDBYPGSPDLVYEVYAGNQYRVAMGADYEELQIPQGSRI 234

RESULT 12
US-10-425-114-71459
; Sequence 71459, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jindong
```



```

QY      165  SSSIEAPLTATDTAIIIS-GSISTONGNCG 192
          |  :  ::  ||  ||  ||  ||  ||  ||
Db      150  --SFGSGFSSFDGTGFTSFGSL----GHGG 172

```

### RESULT 15

```

US-10-369-493-47
: Sequence 47, Application US/10369493
: Publication No. US20030233675A1
: GENERAL INFORMATION:
: APPLICANT: Cao, Yongwei
: APPLICANT: Hinkle, Gregory J.
: APPLICANT: Slater, Steven C.
: APPLICANT: Goldman, Barry S.
: APPLICANT: Chen, Xianfeng
: TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
: FILE REFERENCE: 38-10(52052)B
: CURRENT APPLICATION NUMBER: US/10/369,493
: PRIOR FILING DATE: 2003-02-28
: PRIOR APPLICATION NUMBER: US 60/360,039
: PRIOR FILING DATE: 2002-02-21
: NUMBER OF SEQ ID NOS: 47374
: SEQ ID NO 47
: LENGTH: 376
: TYPE: PR1
: ORGANISM: Agatifex aeolicus
US-10-369-493-47

```

Query Match	7.8%	Score 222.5;	DB 15;	Length 376;
Best Local Similarity	30.4%	Score 432.10		

```

QY      13 EDVYSLNVRREASSEEELKAAVYRLICSMYHPKHDPBELKSAEERLFYVHOAYVLSDP  72
Db      7 KDVEILGVPPNNAOSEELIKAVYRLVVRKXHPDICKPE---CEEKFEKINAAVYLSDP  62

QY      73 QTRAIYDIYKRGKLEMEG-WEVEVERRRP--AEIRBEF-----EQLOREEERRLQORT  123
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      63 EKRKLIDYWGHAAFEGAGAOQRVETTEIPIREILREEFFDPIGSIFERATRRARRRR  122

QY      124 NPKG-TISVGVDATPLPDRIDEYEVDSGSSSPFOLEINKMHISOSIEAPLTITDPAIISG  182
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      123 SVKSEDIYVAPPEIT-----LEBAFK--GTVVP-LEVER-----EVPSCASC-----  159

QY      183 SLSTONGNGGGSINFALRRVTSAKMGELFEFGADLOOLPFLGLKLFRLNLTPLRCFVTNNCA  242
Db      160 -----GCTGYDESKSRCTCPGCGRGE-----TVQGNWF--FOVRQTCPTC-----  197

QY      243 LQFSRRGIRPGLTIVLANLDKNT-VGYLQWRMGISQAMNYSIVBDTKTSHFTVALOLGI  301
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      198 -----GGEQVIYENCHACSTORGY-----GLVNET-----IYXKI  226

QY      302 PHSFALISYQHKFODDDQTRVYKSLKAGPFG-----TVEYGAERKISR-----  345
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      227 PRG-----VRDGSKLVEGSKHAGRYGGRPGDLYIIVKVKPHKIFERKGDLLYDV  277

QY      346 -----HSVIGAANVSIGVPRQSVSLKTKL  367
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      278 NITYPEALVIGTEVEVPTLIDGEEKVKKI  304

```

Search completed: July 13, 2004, 08:52:42  
Job time : 354 secs

***This Page Blank (uspto)***

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 13, 2004, 08:39:16 ; Search time 27 Seconds  
(without alignments)  
1991.521 Million cell updates/sec

Title: US-09-787-678A-6  
Perfect score: 2857  
Sequence: 1 NATALSEELDNEDYSLN.....VLDSALRIPKQSHRIDTDG 559

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_78:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	646	22.6	591	2 T29974	hypothetical prote
2	607.5	21.3	539	2 S58287	J-domain protein D
3	606	21.2	537	2 B84772	probable DnaJ prot
4	437.5	15.3	642	2 T41504	hypothetical DnaJ
5	239.5	8.4	369	2 B72327	dnaJ protein - The
6	235.5	8.2	328	2 G71610	protein with DnaJ
7	224.5	7.9	349	2 B86237	protein F14N23.23
8	222.5	7.8	376	2 E70361	chaperone DnaJ - A
9	218	7.6	354	2 F71623	protein with DnaJ
10	211	7.4	337	2 G84590	probable heat shoc
11	205.5	7.2	350	2 T06689	heat shock protein
12	205	7.2	242	2 JC7933	spermatogenic cell
13	204	7.1	348	2 T04618	heat shock protein
14	201	7.0	335	2 T48161	heat shock protein
15	195	6.8	392	2 B86495	heat shock protein
16	195	6.8	392	2 G72128	heat shock protein
17	194	6.8	340	2 JN0912	heat-shock protein
18	194	6.8	376	2 AH2111	DnaJ protein [limp
19	192.5	6.7	331	2 A86621	probable heat shoc
20	192.5	6.7	418	2 S42031	LDJ2 protein - lee
21	192	6.7	178	2 G02272	heat shock protein
22	192	6.7	420	2 T49127	dnaJ protein homol
23	192	6.7	420	2 S71199	dnaJ protein homol
24	190	6.6	389	2 A64202	heat shock protein
25	189	6.6	390	2 S73459	heat shock protein
26	188.5	6.6	423	2 T09338	DnaJ-like protein
27	187.5	6.6	373	2 DB1242	dnaJ protein homol
28	187.5	6.6	419	2 T07371	dnaJ protein homol
29	187.5	6.6	499	2 G96831	hypothetical prote

30	187	6.5	351	2 S23508	dnaJ protein homol
31	185.5	6.5	355	2 T40385	hypothetical prote
32	185.5	6.5	910	1 S73361	dnaJ homolog
33	185	6.5	377	2 A83052	DnaJ protein PA476
34	184.5	6.5	369	2 D64686	co-chaperone and h
35	184.5	6.5	391	2 G84611	probable DnaJ prot
36	183.5	6.4	346	2 T48049	hypothetical prote
37	182	6.4	336	2 T09133	heat shock protein
38	182	6.4	373	2 F81333	chaperone DnaJ Cj1
39	181	6.3	417	2 UQ2142	chaperone ANU1 pro
40	180.5	6.3	332	2 S76622	hypothetical prote
41	180.5	6.3	376	2 G90630	DnaJ protein [limp
42	180.5	6.3	376	2 G85481	chaperone with Dna
43	180.5	6.3	392	2 H71526	probable heat shoc
44	180.5	6.3	392	2 DB1693	dnaJ protein homol
45	180.5	6.3	419	2 T01643	DnaJ protein homol

ALIGNMENTS

RESULT 1	
T29974	hypothetical protein F11G11.7 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans	
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999	
C:Accession: T29974	
R:Laraille, P.; Deadman, R.	
A:Submitted to the EMBL Data Library, November 1996	
A:Description: The sequence of C. elegans cosmid F11G11.	
A:Reference number: Z20715	
A:Accession: T29974	
A:Status: preliminary; translated from GB/EMBL/DBJ	
A:Molecule type: DNA	
A:Residues: 1-591 <LAT>	
A:Cross-references: EMBL:U80451; PIDN:AA37835.1; GSPDB:GN00020; CESP:F11G11.7	
A:Experimental source: strain Bristol N2; clone F11G11	
C:Genetics:	
A:Gene: CESP:F11G11.7	
A:Map position: 2	
A:Introns: 36/3; 213/3; 288/1; 541/3; 585/1	
Query Match	22.6%; Score 646; DB 2; Length 591;
Best local Similarity	30.2%; Pred. No. 6.7e-39;
Matches 181; Conservative 110; Mismatches 231; Indels 78; Gaps 19;	
QY	7 EEBLDNEDYSLNVRREASSEEIKAYRRLCMYHPDKRDDELKSOAERLFTNLVHOAY 66
DB	22 EEBI---DFAILNVPRKATDDDEIKAYRRCIMFPHDRVNDKKDAERVFVKLRRAH 78
QY	67 EVLSDPQTRAIYDIYKRGKLEMEGWEVERRRTPAETREEFERLQREERERRLOQRTNPK 126
DB	79 EVLIDPKQRAIYVALGQGLDTQGWELVSRNAPENIRKEVEFLQIKDELMLQORVHPT 138
QY	127 GTISGVDAATDLDPRYDEBYEDVSGSFPQIEINIKHMSQIEAPLPAATATLISGSLST 186
DB	139 SAFWIKTTIGMFQENDE-----RYPP-QLLGSLSSVDCATGTVDRFGLSERVKTI 190
QY	187 QNGNGGSSINFALRYVTSAGKMGLEFGADLQGPL-----FGIKLFRNLTPRCFVT 238
DB	191 GNGRGDSVSAVMKVA-----GSVNLNNTISLSAESVSLTCRAARNVFTTRAIVI 240
QY	239 TNCALQPS--SRGIRGLTTLVLANIDK-----NTVGY---LQWR--- 274
DB	241 VQPOLQYNMLHEAIIPIISIAMSMNRIKKVYSEKDRLEFTDSIFCKIPFYSMRHTTRQGS 300
QY	275 -----GIOSAMNTSYRDTKTSHTFVALQL---GIHPSPALSYQHKFPDDQOTRYKGS 325
DB	301 IVNLISPIANALTTTWH--TENNHAQAVGSLTSPINSVRLVYTRKPENDSITTEW--S 357
QY	326 LKAGFGCTVVEGAERKIRSHSVLGAASVGVQD--VSLKVKLNRAISQTYFPPIHLLTD-- 382
DB	358 VQLTGYGVNPAIMDRISRSRISCSFHSFSPCLLYTKFKLAKAGSTFDWQIVLDCDK 417





C:Species: Schizosaccharomyces pombe  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999  
C:Accession: T41504  
R:Murphy, L.; Harris, D.; Wood, V.; Rajandream, M.A.; Barrell, B.G.  
submitted to the EMBL Data Library, March 1999  
A:Reference number: 221999  
A:Accession: T41504  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-642 <MOL>  
A:Cross-references: EMBL:AL049522; PIDN:CAB40007.1; GSPDB:GN00068; SPDB:SPCC63.03  
C:Genetics:  
A:Experimental source: strain 972h-; cosmid c63  
A:Gene: SPDB:SPCC63.03  
A:Map position: 3  
A:introns: 18/3; 423/2; 563/3

Query Match 15.3%; Score 437.5; DB 2; Length 642;  
Best Local Similarity 24.0%; Pred. No. 4.5e-23;  
Matches 151; Conservative 120; Mismatches 246; Indels 111; Gaps 20;

QY 8 EELNEDYYSLLNVRASSEELKAAYRLCMLYHPDKRDELKSGAERLFNLVQAYE 67  
DB 2 DEADSEWLYLALGLPKDATSDQIKESYRLRLPHPDH--TADQAAAEKRFQIIQAYE 60  
QY 68 VLSDPQTALIDYIKRGLMEGWVEYERRTPAEIRFEERFLQREERERLQORTNPKG 127  
DB 61 VLSDPKKEIYDNFGEQGLKTD--NMGVFPKGSABELKKKIREQIOERDIHRIDSLVQSR 119  
QY 128 TISVGVDTDLFDRDEVEDVSG-----SSPQIENNMHSISLEAPLATD 176  
DB 120 ETTIVNMTPLFAR--NIVQNALGLGAGTRMLTPYERPSLIQWVSFOIKSSFSIPTSPSN 178  
QY 177 -----TALLSGSLSTO-----NGNGGGSINFALRYTSAKMGEL----- 211  
DB 179 DLKRFSPFSFGSGDDEDFSAPSDEDEGNHNTSSSLSTVTEASMRQNSKLOPSIFAVYHS 238  
QY 212 -----EFGAGDIQGLPLFGKLFRNLTPRCFVTTNCALPSSRGIRPGLITVLAARLD 263  
DB 239 QPSPNLSSEIGFSLRPGLITVKSVAIINNQTFIVP--LIQISGLKRPQATVIGRQIT 296  
QY 264 KNTGVYLQMR-----WGIOAMNTSIVADTKTSHTVLAQ--LGIHPS----- 304  
DB 297 RFGITLAAWKQGVWSLGSWGIASPRGAN-----SSSLSLQQOKALPNSLVPLQSLMNA 349  
QY 305 -----FALISYQHKFQ---DDQTRVK--GSLKAGFGTVEYEAERKISRHSVLGAAY 353  
DB 350 EYTAGLWMSGLAIVNNLNKATNEDSPYQIKLGTSMTVGGLQVSDGTSKVKGRYSTPFGNI 409  
QY 354 SVGVAPQG--VSLKVKLNRAISQTYFFPIHLTDQLPSAMFYATVGLVYVYFAMRLLIKP-- 410  
DB 410 SVGVAPGTSITFSILNMSRGLQKISLPIWMSVFDSSAVFWGLVFPITSLIGVEQFLRRR 469  
QY 411 -----YLAQKEKELEKQRESAATDVLOKKQAEASAVPLMDQSPRRILAEASRGL 462  
DB 470 LSNQRLRLRLRQIKDKDSQE-----RRKYSAIRAVLMEIYEVKKQKLEMEKGL 519  
QY 463 IIVNAMYKFPVNDKSRKSEKYVLDVTPLOCLVDSKL--LTTEASKAGLPGFYDPCVGE 521  
DB 520 VIEYAEY--RVVNCANEDPKQ--DVTISIALVENSRLAIPSSVKSISIIIGIYPLPSDN 576  
QY 522 EKNILKVLQVFRGVLIHQVWVLDSEALRIPKQ 551  
DB 577 EKLEIYVTFHQQRHVVLRDQGVFLPSR 606

RESULT 5  
B72327  
dnaJ proteain - Thermotoga maritima (strain MSB8)  
C:Species: Thermotoga maritima  
C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
C:Accession: B72327  
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey

C.M. Garrett, M.M., Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.

Nature 399, 323-329, 1999

A>Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome se

A:Reference number: A72200; MUID:99287316; PMID:10360571

A:Accession: B72327

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-369 <ARN>

A:Cross-references: GB:AE001751; GB:AE000512; NID:G4981371; PIDN:AA035931.1; PID:G49813

A:Experimental source: strain MSB8

C:Genetics:

A:Gene: TM0849

C:Superfamily: heat shock protein dnaJ; dnaJ amino-terminal homology

F:7-73/Domain: dnaJ amino-terminal homology <DNU>

Query Match 8.4%; Score 239.5; DB 2; Length 369;

Best Local Similarity 22.5%; Pred. No.2.2e-09;

Matches 103; Conservative 62; Mismatches 153; Indels 139; Gaps 17;

QY 11 DNEDYSLNLYRREASSELKAAYRRLCMLYHPDKHRDPKLSQAEKRLFNLVHQAYEVL 70

DB 4 EKQDYELLGVPDATOGEIKRAYKRLVKNHMPDRH--PENRKEAQRFKEIQEAYEVL 61

QY 71 DPQRAITDIYKKGKLEMEGWEVVERRTPAIREEPRRL-----QGRERRR 118

DB 62 DPQRAAMYDRFGYVG-EQPTVQETESGGFPDDIFDFENINRDI FDVFFGGERPHQEBRR 120

QY 119 LQQRNPKGTISVGADATDLPDRYDEYEDVSGSFPOIEINKHISQSIAPLATDTA 178

DB 121 EYARGEIRIREIVTSLD-----INGAEIP-VVERY- 153

QY 179 ILSGSLSTONGNGGGSINPALRVYTSAKWGELEFGAGDLQGLFGLKLFNLTGRCVYT 238

DB 154 ---ETCRGCGTGVFPNAGYMDGPS CGGTGRIR---EERRSPFGYFVSERTCERCGGT 205

QY 239 TNCALQF-----SSRGIRPGLTIVLANLDKATVGYLQMRKGIQSAWNTSYVRPTKISH 292

DB 206 GKIPREYCHECGSGGRVLRK-----VRRT--- 229

QY 293 FTVALLQGIHPSPALISYQHKFODDQTRVYGLSKAGPFG-----TVVEXGARKKISR 345

DB 230 -----VKIPRVEDGTHLRITGGGNAGYGGPYGDLITIVRYKPPRRFK 274

QY 346 -----HSVLGAASVGVPOG--VSLKVKLNRASTQTFEPIHLTDQLPSAM 389

DB 275 SGGSLDVYDVTIDYQALIGTTVEVPLPREGTVMKIKIPGTQPEYF---RLKGKGLPR- 330

QY 390 FYATVGPLVVFYFAMHRLIKPYLPAQKE--KELEKOR 424

DB 331 -YGRRGDLIV--NVHVEIPKSLSREERKYLEELAKR 364

RESULT 6

G71610

protein with DnaJ domain, DNU1/SIS1 family PFB0595W - malaria parasite (Plasmodium falc

C:Species: Plasmodium falciparum

C:Date: 13-Nov-1998 #sequence\_revision 13-Nov-1998 #text\_change 21-Jul-2000

C:Accession: G71610

R:Gardner, M.J.; Tetteelin, H.; Carnucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.

I: Perles, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.

Science 282, 1126-1132, 1998

A>Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.

A:Reference number: A71600; MUID:99021743; PMID:9804551

A:Accession: G71610

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-328 <GAR>

A:Cross-references: GB:AE001405; GB:AE001362; NID:G3845224; PIDN:AA071908.1; PID:G38452

A:Experimental source: clone 3D7

C:Genetics:

A:Gene: PFB0595W

C:Superfamily: heat shock protein dnaJ; dnaJ amino-terminal homology

F/4-72/Domain: dnaJ amino-terminal homology <DNJ>

Query Match 8.2%; Score 235.5; DB 2; Length 328;  
Best Local Similarity 24.9%; Pred. No. 3.6e-09;  
Matches 64; Conservative 38; Mismatches 64; Indels 91; Gaps 7;  
13 EDVYSLNVRREASSELKAAVRLCMLYHPDKHDPDELKSOAERLFNLVHQAYEVLSDP 72  
3 KDYYSILGVRDCTNDTKKAYRKLMWHPDKHNDKSKCKAEKFKNAEAYDLADE 62  
73 QTRAIYDIYGRKGLMEGWVERRRRTPAERIEEFERLQREERERLQORTNKGTISV 131  
63 EKRIYDTYGEGL-----KGSIPNG 83  
132 -----GVDTATLDFR-----YDEYEDVS-----GSSFPQLEIN 160  
84 GNTYVYSGVPSSELSFISDSQFSTSTFDEDFSPSTFVMTSRKSPSTTTNINT 143  
161 KMHISQSIAPLTATDTAILSGS-----LSTONGGGSINFALRYTSAGWG 209  
144 NKKPRTYEVPLSLSEELSGCKKKLTKRKRFMGTKSYEDNNTYIDVK-----AGMK 198  
210 E----LEFGAGDIQGPL 222  
199 DGTKITFYGGEDQLSPM 215

## RESULT 7

protein F14N23.23 [imported] - Arabidopsis thaliana  
C/Species: Arabidopsis thaliana (mouse-ear cress)  
C/Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C/Accession: E86237  
R/Theologas, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chn, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
Nansen, N.F.; Hughes, B.; Hultz, L.  
Nature 408, 816-820, 2000  
A/Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luo, J.S.; Maiti, R.; Marziani,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.; Schatz, G.; Shim, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A/Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A/Reference number: A86141; MUID:21016719; PMID:11130712  
A/Accession: E86237  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-349 <STO>  
A/Cross-references: GB:AE005172; NID:G4914337; PIDN:AAD32885.1; GSPDB:GN00141  
C/Genetics:  
A/Gene: F14N23.23  
A/Map position: 1  
C/Superfamily: heat shock protein dnaJ; dnaJ amino-terminal homology

Query Match 7.9%; Score 224.5; DB 2; Length 349;  
Best Local Similarity 31.1%; Pred. No. 2.4e-08;  
Matches 60; Conservative 22; Mismatches 62; Indels 49; Gaps 5;  
14 DYVSLNVRREASSELKAAVRLCMLYHPDKHDPDELKSOAERLFNLVHQAYEVLSDP 73  
4 DYVNVKVNANEDDKSKYRRMAKWHDPK--NPSKKEAKFKFOISEADVLSDP 61  
74 TRAIYDIYGRKGLMEGWVERRRRTPAERIEEFERLQREERERLQORTNKGTISV 133  
62 RRIYDYGEGLSTD-----LPTAETRAHQOQRSSNSEFRYPR----- 106  
134 DATDLFDRYDEYEDVSGSFPQLEINMHISQSIAPLTATDTAILSGSLSTONGGGS 193  
107 DAEDFAEFEGEGDAFGG-----SSGRTREDGDDGG 140  
194 SINPALRYTSAG 206  
141 -----RRFKSAE 147

## RESULT 8

E70361  
chaperone DnaJ - Aquifex aeolicus  
C/Species: Aquifex aeolicus  
C/Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 20-Aug-1999  
C/Accession: E70361  
R/Beckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; V.  
Nature 392, 353-358, 1998  
A/Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.  
A/Reference number: A70300; MUID:98196666; PMID:9537320  
A/Accession: E70361  
A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
A/Molecule type: DNA  
A/Residues: 1-376 <AQF>  
A/Cross-references: GB:AE00703; NID:G2983287; PIDN:AAC06881.1; PID:G2983289; GB:AE00065  
C/Genetics:  
A/Gene: dnaJ2  
C/Superfamily: heat shock protein dnaJ; dnaJ amino-terminal homology  
F/8-72/Domain: dnaJ amino-terminal homology <DNJ>

Query Match 7.8%; Score 222.5; DB 2; Length 376;  
Best Local Similarity 26.4%; Pred. No. 3.6e-08;  
Matches 102; Conservative 44; Mismatches 120; Indels 121; Gaps 19;

13 EDVYSLNVRREASSELKAAVRLCMLYHPDKHDPDELKSOAERLFNLVHQAYEVLSDP 72  
7 KDYSELGVPRNAGSEIKKAYRLVRKYNPDIKKPE-----CEEKFEINBAVQLSDP 62  
73 QTRAIYDIYGRKGLMEGWVERRRRT- AEIEEF-----ERLQREERERLQORT 123  
63 EKKRYIDMGHAIFGAGAQOQVETTELPIREILREFDFDGSIFERATGRRAARRR 122  
124 NPKG-TISGVDTATLDFRYDEYEDVSGSFPQLEINMHISQSIAPLTATDTAILSG 182  
123 SYKGEIVVPVIT-----LEAFK---CTYTP-IEVER-----EYGCSCAC----- 159  
183 SLSTONGGGSINFALRYTSAGWGELEFGAGDIQGPLFKLFRNLTPRCFVTTNCA 242  
160 -----GGTGYDESKSRFCPCGGRGE-----TVQGNWF--FQVRQCPCPC----- 197  
243 LQSSKGRIPGLTYLAARLDKNT--VGYQMWGQSMNNTIVDDTKSHFTVALQGI 301  
198 -----GEGVAYENCHACTGRGY-----GLVRET-----IKVKI 226  
302 PHSFALISYQHKFODDDQTRVKSGLKAGFFG-----TVVEYGAERKISR----- 345  
227 PPG-----VRGSKLVVBGKGAAGRGPPDDLYIIVAKRHKIFERKGDLLYDV 277  
346 -----HSVIGAAVSVGVPOGSLKVKL 367  
278 NITYPEAVLVGEVEVPTLDGEKVKVKI 304

## RESULT 9

F71623  
protein with DnaJ domain PFB0090C - malaria parasite (Plasmodium falciparum)  
C/Species: Plasmodium falciparum  
C/Date: 13-Nov-1998 #sequence\_revision 13-Nov-1998 #text\_change 21-Jul-2000  
C/Accession: F71623  
R/Gardner, M.J.; Tectelin, H.; Carnucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;  
Petter, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.  
Science 282, 1126-1132, 1998  
A/Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.  
A/Reference number: A71600; MUID:99021743; PMID:9804551  
A/Accession: F71623  
A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
A/Molecule type: DNA  
A/Residues: 1-354 <GAR>  
A/Cross-references: GB:AE001370; GB:AE001362; NID:G3845089; PIDN:AAC71808.1; PID:G384509





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 13, 2004, 08:39:16 ; Search time 75 Seconds

(without alignments)  
2351.662 Million cell updates/sec

Title: US-09-787-678A-6

Perfect score: 2857

Sequence: 1 MATLSEEDLNDDYSLN.....VLDEALRIPKSHRIDTG 559

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

1017041

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_virus:\*
- 16: sp\_bacteriap:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARY

Result No.	Score	Query Match	Length	ID	Description
1	2848	99.7	559	4	Q9NVH1
2	2826	98.9	552	4	Q8NDM2
3	2792	97.7	559	11	Q8CGU5
4	2552	89.3	507	4	Q96CJ7
5	2461	86.1	563	13	Q802X9
6	1677	58.7	327	11	Q8C1Z4
7	1452	50.8	330	13	Q7SZY3
8	1220	42.7	318	4	Q9VMU8
9	1160.5	40.6	545	5	Q9V6Z8
10	1022	35.8	214	11	Q8BP83
11	661	23.1	569	10	Q8LN96
12	646	22.6	591	5	P91243
13	607.5	21.3	539	10	Q39079
14	606	21.2	537	10	Q9SL89
15	437.5	15.3	642	3	Q9Y7T0
16	363	12.7	82	4	Q7Z6Y9

17	311	10.9	79	13	Q7SZY4	Q7SZY4 brachydanio
18	235.5	8.2	328	5	Q96212	Q96212 plasmodium
19	225	8.0	401	5	Q96123	Q96123 plasmodium
20	226.5	7.9	349	10	Q8LAU1	Q8LAU1 arabidopsis
21	224.5	7.9	349	10	Q9SY77	Q9SY77 arabidopsis
22	224	7.8	242	6	Q8WN90	Q8WN90 bos taurus
23	220.5	7.7	342	10	Q84SD3	Q84SD3 oryza sativ
24	219.5	7.7	145	4	Q8WRF6	Q8WRF6 homo sapien
25	217	7.6	312	11	Q9DN41	Q9DN41 mus musculu
26	212	7.4	433	5	Q81489	Q81489 plasmodium
27	211	7.4	337	10	Q9SIL3	Q9SIL3 arabidopsis
28	207	7.2	261	11	Q9DAN5	Q9DAN5 mus musculu
29	205.5	7.2	350	10	Q9SUS7	Q9SUS7 arabidopsis
30	205	7.2	242	6	Q86224	Q86224 macaca fasc
31	204	7.1	348	10	Q49457	Q49457 arabidopsis
32	203	7.1	323	10	Q8L8S5	Q8L8S5 arabidopsis
33	203	7.1	323	10	Q9SR91	Q9SR91 arabidopsis
34	203	7.1	515	10	Q8H812	Q8H812 oryza sativ
35	201	7.0	335	10	Q9M034	Q9M034 arabidopsis
36	200.5	7.0	237	13	Q7ZM83	Q7ZM83 brachydanio
37	196	6.9	273	5	Q7YTD1	Q7YTD1 cryptospori
38	195.5	6.8	424	5	Q8IL88	Q8IL88 plasmodium
39	195	6.8	385	16	Q7VG06	Q7VG06 helicobacte
40	194	6.8	342	10	Q9PEW7	Q9PEW7 lycopersico
41	194	6.8	376	16	Q8YU45	Q8YU45 anabaena sp
42	193.5	6.8	259	5	Q9V7K6	Q9V7K6 drosophila
43	193	6.8	375	16	Q8DKR7	Q8DKR7 synchococc
44	193	6.8	388	16	Q8EM04	Q8EM04 mycoplasma
45	192.5	6.7	331	10	Q9XIF5	Q9XIF5 arabidopsis

#### ALIGNMENTS

#### RESULT 1

Q9NVH1 PRELIMINARY; PRT; 559 AA.

AC Q9NVH1; 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
DE Hypothetical protein FLJ10737.  
OS Homo sapiens (Human)  
OC Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_Taxid=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Isegal T., Oca T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,  
RA Watanabe M., Hosoi T., Kaku Y., Kodaira H., Kondo Y., Sugawara M.,  
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,  
RA Watanabe S., Kimura K., Murakami K., Iehi S., Kawai Y., Saito K.,  
RA Yamamoto J., Kakumaru A., Nakamura Y., Nagahari K., Masuno Y.,  
RA Minomura K., Iwayanagi T.,  
RT "NEDO human cDNA sequencing project";  
RL Submitted (FEb-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AK001599; BAA91780.1; -  
DR HSBP; P25685; IHDJ.  
DR InterPro: IPR001623; DnaJ\_N.  
DR Pfam; PF00226; DnaJ\_1.  
DR SMART; SM00271; DnaJ\_1.  
DR PROSITE; PS00636; DnaJ\_1; 1.  
DR PROSITE; PSS0076; DnaJ\_2; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 559 AA; 63336 MW; A610F997833276A0 CRC64;

Query Match 99.7%; Score 2848; DB 4; Length 559;

Best Local Similarity 99.6%; Pred. No. 2.3e-211;

Matches 557; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```
Db 1 MATALSEELDNEDYSSLNVRREASSEELKAAVRLCMLYHPDKRDELKSOAERLFN 60
Qy 61 LVHQAAYVLSDPQTRAIYDIYGRKGLMEGWVERRRTPAEIREEERLQRRERRLQ 120
Db 61 LVHQAAYVLSDPQTRAIYDIYGRKGLMEGWVERRRTPAEIREEERLQRRERRLQ 120
Qy 121 QRTNPKKTIISVGDATDLFDRYDEEYEDVSGSFPQIEINKMHSQSIAPLATATTAI 180
Db 121 QRTNPKKTIISVGDATDLFDRYDEEYEDVSGSFPQIEINKMHSQSIAPLATATTAI 180
Qy 181 SGSLSTONGGGSINPALRRVTSAKGWELERFAGDLOGPLFGLKLFNNLTPRCVTTN 240
Db 181 SGSLSTONGGGSINPALRRVTSAKGWELERFAGDLOGPLFGLKLFNNLTPRCVTTN 240
Qy 241 CALQFSRGIRPGLITVLARNLDKNVTGYLQWRMGIOSAMNTSIYVDTKTSHTVLAOLG 300
Db 241 CALQFSRGIRPGLITVLARNLDKNVTGYLQWRMGIOSAMNTSIYVDTKTSHTVLAOLG 300
Qy 301 IPHSFALISYOHKFQDDQTRVKSILKAGFPGTVVEYGAERKISRHSVLGAASVGPQG 360
Db 301 IPHSFALISYOHKFQDDQTRVKSILKAGFPGTVVEYGAERKISRHSVLGAASVGPQG 360
Qy 361 VSLKVKLNRASTQYFFPIHLTDOLPSAMFYATVGPLVYFAMHRLIIPYLAQKEKEL 420
Db 361 VSLKVKLNRASTQYFFPIHLTDOLPSAMFYATVGPLVYFAMHRLIIPYLAQKEKEL 420
Qy 421 EKQRESATDVLOKKQASAVRLMOSVRRITIEAESRMGLIIVAMWCKFVNDKSRKS 480
Db 421 EKQRESATDVLOKKQASAVRLMOSVRRITIEAESRMGLIIVAMWCKFVNDKSRKS 480
Qy 481 EKVAVIDVTVPLQCLVKSILITEASKAGLPGFYDPCVGEENKLVLYQFRGVLHQWV 540
Db 481 EKVAVIDVTVPLQCLVKSILITEASKAGLPGFYDPCVGEENKLVLYQFRGVLHQWV 540
Qy 541 LDSEBALRIPKQSHRIDTDG 559
Db 541 LDSEBALRIPKQSHRIDTDG 559
```

## RESULT 2

```
Q8NDM2 ID Q8NDM2 PRELIMINARY; PRT; 552 AA.
AC Q8NDM2;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein (Fragment).
GN DKFZP727C181.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Breast cancer;
RA Pouscka A., Klein M., Mewes H.W., Gassenhuber J., Wiemann S.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL833841, CAD38701.1, -.
DR InterPro: IPR001623, DnaJ_N.
DR Pfam: PF00226, DnaJ_1.
DR SMART: SM00271, DnaJ_1.
DR PROSITE: PS00636, DnaJ_1; 1.
DR PROSITE: PS50076, DnaJ_2; 1.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 552 AA; 62574 MW; 085A3748F5285C0C CRC64;
```

```
Query Match 98.9%; Score 2826; DB 4; Length 552;
Best Local Similarity 100.0%; Pred. No. 1,1e-209;
Matches 552; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
8 EELDNEDYSSLNVRREASSEELKAAVRLCMLYHPDKRDELKSOAERLFNLVHQAAYE 67
```

```
Db 1 EELDNEDYSSLNVRREASSEELKAAVRLCMLYHPDKRDELKSOAERLFNLVHQAAYE 60
Qy 68 VLSDPQTRAIYDIYGRKGLMEGWVERRRTPAEIREEERLQRRERRLQRRERRLQ 127
Db 61 VLSDPQTRAIYDIYGRKGLMEGWVERRRTPAEIREEERLQRRERRLQRRERRLQ 120
Qy 128 TTSVGDATDLFDRYDEEYEDVSGSFPQIEINKMHSQSIAPLATATTAIISGLSTQ 187
Db 121 TTSVGDATDLFDRYDEEYEDVSGSFPQIEINKMHSQSIAPLATATTAIISGLSTQ 180
Qy 188 NNGGGSINPALRRVTSAKGWELERFAGDLOGPLFGLKLFNNLTPRCVTTNCAIQFS 247
Db 181 NNGGGSINPALRRVTSAKGWELERFAGDLOGPLFGLKLFNNLTPRCVTTNCAIQFS 240
Qy 248 RGRPELITVLARNLDKNVTGYLQWRMGIOSAMNTSIYVDTKTSHTVLAOLGIPHSFL 307
Db 241 RGRPELITVLARNLDKNVTGYLQWRMGIOSAMNTSIYVDTKTSHTVLAOLGIPHSFL 300
Qy 308 ISYOHKFQDDQTRVKSILKAGFPGTVVEYGAERKISRHSVLGAASVGPQVSLKVL 367
Db 301 ISYOHKFQDDQTRVKSILKAGFPGTVVEYGAERKISRHSVLGAASVGPQVSLKVL 360
Qy 368 NRASQTYFFPIHLTDOLPSAMFYATVGPLVYFAMHRLIIPYLAQKEKELKQRESA 427
Db 361 NRASQTYFFPIHLTDOLPSAMFYATVGPLVYFAMHRLIIPYLAQKEKELKQRESA 420
Qy 428 ATDVLOKKQASAVRLMOSVRRITIEAESRMGLIIVAMWCKFVNDKSRSEKVKVD 487
Db 421 ATDVLOKKQASAVRLMOSVRRITIEAESRMGLIIVAMWCKFVNDKSRSEKVKVD 480
Qy 488 VTVPLQCLVKSILITEASKAGLPGFYDPCVGEENKLVLYQFRGVLHQWVLDSEALR 547
Db 481 VTVPLQCLVKSILITEASKAGLPGFYDPCVGEENKLVLYQFRGVLHQWVLDSEALR 540
Qy 548 IPKQSHRIDTDG 559
Db 541 IPKQSHRIDTDG 552
```

## RESULT 3

```
Q8C6U5 ID Q8C6U5 PRELIMINARY; PRT; 559 AA.
AC Q8C6U5;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical DnaJ N-terminal domain containing protein.
GN E030019A03RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Lung;
RX MEDLINE=22354683; PubMed=12466851;
RA The RANOM Consortium,
RL "Analysis of the mouse transcriptome based on functional annotation of
RT the RIKEN Genome Exploration Research Group Phase I & II Team;
RT 60,770 full-length cDNAs."
RL Nature 420:563-573 (2002).
DR EMBL: AK053156, BAC35287.1, -.
DR MGD: MGI:2443386, E030019A03RIK.
DR InterPro: IPR001623, DnaJ_N.
DR Pfam: PF00226, DnaJ_1.
DR SMART: SM00271, DnaJ_1.
DR PROSITE: PS00636, DnaJ_1; 1.
DR PROSITE: PS50076, DnaJ_2; 1.
KW Hypothetical protein.
SQ SEQUENCE 559 AA; 63233 MW; DA6B7590EC7D2561 CRC64;
```

```
Query Match 97.7%; Score 2792; DB 11; Length 559;
Best Local Similarity 96.6%; Pred. No. 4.9e-207;
```

Matches 540; Conservative 14; Mismatches 5; Indels 0; Gaps 0;

```
QY 1 MATALSEELDNEDYYSLNVRREASSEELKAAVRLCMLYHPDKHRDPDLKSOAERLNF 60
D 1 MATALSEELDNEDYYSLNVRREASSEELKAAVRLCMLYHPDKHRDPDLKSOAERLNF 60
QY 61 LVHQAAYVLSDPOTRAIYDIYGRKGLMEGMEVVERRRTPAIREEPERLQREBERRLQ 120
D 61 LVHQAAYVLSDPOTRAIYDIYGRKGLMEGMEVVERRRTPAIREEPERLQREBERRLQ 120
QY 121 QRTNPKGTISVGVDATDLFDRYDEYEDVSGSSFPQIEINKMHISQSIAPLATDTAIL 180
D 121 QRTNPKGTISVGVDATDLFDRYDEYEDVSGSSFPQIEINKMHISQSIAPLATDTAIL 180
QY 181 SGSLSTONGGSSINPALRVTSAGKMGELFEGADLQGPLFGLKLFRLTTRCFVTNN 240
D 181 SGSLSTONGGSSINPALRVTSAGKMGELFEGADLQGPLFGLKLFRLTTRCFVTNN 240
QY 241 CALQFSRGRIRPGITTYLAANLDKNTVGYLQMRNGIOSAMNTSIVRDTKTSHTFVALQLG 300
D 241 CALQFSRGRIRPGITTYLAANLDKNTVGYLQMRNGIOSAMNTSIVRDTKTSHTFVALQLG 300
QY 301 IPHSFALISYQHKFODDQTRVKSLKAGFGTIVEYGAERKISRHSVLGAASVGVPOG 360
D 301 IPHSFALISYQHKFODDQTRVKSLKAGFGTIVEYGAERKISRHSVLGAASVGVPOG 360
QY 361 VSLKVKLNRRASQTYFFPIHLTDQLLPSAMFYATGVLVYFAMHRLIKPYLBAQKEKEL 420
D 361 VSLKVKLNRRASQTYFFPIHLTDQLLPSAMFYATGVLVYFAMHRLIKPYLBAQKEKEL 420
QY 421 EKORESATDVLQKKOEAESAVALMOESVRIIEAESRMGLIIVNAMYGFVNDKSRKS 480
D 421 EKORESATDVLQKKOEAESAVALMOESVRIIEAESRMGLIIVNAMYGFVNDKSRKS 480
QY 481 EKXKVIDVTVPLOCLVXDSKILITEASKAGLPFGYDPCVGEKXNLKVLVQFRGLHQVMV 540
D 481 EKXKVIDVTVPLOCLVXDSKILITEASKAGLPFGYDPCVGEKXNLKVLVQFRGLHQVMV 540
QY 541 LDSEALRIPKQSHRIDTDG 559
D 541 LDSEALRIPKQSHRIDTDG 559
```

## RESULT 4

```
QY 096CL7 PRELIMINARY; PRT; 507 AA.
AC 096CL7;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Strauberg R.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC014145; AAH14145.1;
DR InterPro; IPR001623; DnaJ_N.
DR Pfam; PF00226; DnaJ_1.
DR SMART; SM00271; DnaJ_1.
DR PROSITE; PS00636; DnaJ_1;
DR PROSITE; PS50076; DnaJ_2;
KM Hypothetical protein.
SQ SEQUENCE 507 AA; 57196 MW; 72BD08620B512A CRC64;
```

Query Match 89.3%; Score 2552; DB 4; Length 507;  
Best Local Similarity 90.7%; Pred. No. 1.5e-181;  
Matches 507; Conservative 0; Mismatches 0; Indels 52; Gaps 1;

QY 1 MATALSEELDNEDYYSLNVRREASSEELKAAVRLCMLYHPDKHRDPDLKSOAERLNF 60

```
D 1 MATALSEELDNEDYYSLNVRREASSEELKAAVRLCMLYHPDKHRDPDLKSOAERLNF 60
QY 61 LVHQAAYVLSDPOTRAIYDIYGRKGLMEGMEVVERRRTPAIREEPERLQREBERRLQ 120
D 61 LVHQAAYVLSDPOTRAIYDIYGRKGLMEGMEVVERRRTPAIREEPERLQREBERRLQ 120
QY 121 QRTNPKGTISVGVDATDLFDRYDEYEDVSGSSFPQIEINKMHISQSIAPLATDTAIL 180
D 121 QRTNPKGTISVGVDATDLFDRYDEYEDVSGSSFPQIEINKMHISQSIAPLATDTAIL 180
QY 181 SGSLSTONGGSSINPALRVTSAGKMGELFEGADLQGPLFGLKLFRLTTRCFVTNN 240
D 181 SGSLSTONGGSSINPALRVTSAGKMGELFEGADLQGPLFGLKLFRLTTRCFVTNN 240
QY 241 CALQFSRGRIRPGITTYLAANLDKNTVGYLQMRNGIOSAMNTSIVRDTKTSHTFVALQLG 300
D 241 CALQFSRGRIRPGITTYLAANLDKNTVGYLQMRNGIOSAMNTSIVRDTKTSHTFVALQLG 300
QY 301 IPHSFALISYQHKFODDQTRVKSLKAGFGTIVEYGAERKISRHSVLGAASVGVPOG 360
D 301 IPHSFALISYQHKFODDQTRVKSLKAGFGTIVEYGAERKISRHSVLGAASVGVPOG 360
QY 361 VSLKVKLNRRASQTYFFPIHLTDQLLPSAMFYATGVLVYFAMHRLIKPYLBAQKEKEL 420
D 361 VSLKVKLNRRASQTYFFPIHLTDQLLPSAMFYATGVLVYFAMHRLIKPYLBAQKEKEL 420
QY 421 EKORESATDVLQKKOEAESAVALMOESVRIIEAESRMGLIIVNAMYGFVNDKSRKS 480
D 421 EKORESATDVLQKKOEAESAVALMOESVRIIEAESRMGLIIVNAMYGFVNDKSRKS 480
QY 481 EKXKVIDVTVPLOCLVXDSKILITEASKAGLPFGYDPCVGEKXNLKVLVQFRGLHQVMV 540
D 481 EKXKVIDVTVPLOCLVXDSKILITEASKAGLPFGYDPCVGEKXNLKVLVQFRGLHQVMV 540
QY 541 LDSEALRIPKQSHRIDTDG 559
D 541 LDSEALRIPKQSHRIDTDG 559
```

## RESULT 5

```
QY 0802X9 PRELIMINARY; PRT; 563 AA.
AC 0802X9;
DT 01-JUN-2003 (Tremblrel. 24, Created)
DT 01-JUN-2003 (Tremblrel. 24, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Similar to RIKEN CDNA E030019A03 gene.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Strauberg R.;
RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC046897; AAH46897.1;
DR InterPro; IPR001623; DnaJ_N.
DR Pfam; PF00226; DnaJ_1.
DR SMART; SM00271; DnaJ_1.
DR PROSITE; PS00636; DnaJ_1;
DR PROSITE; PS50076; DnaJ_2;
SQ SEQUENCE 563 AA; 63422 MW; B9ECE276CCEB320F CRC64;
```

Query Match 86.1%; Score 2461; DB 13; Length 563;  
Best Local Similarity 82.2%; Pred. No. 1.9e-181;  
Matches 462; Conservative 59; Mismatches 37; Indels 4; Gaps 1;

QY 1 MATALSEELDNEDYYSLNVRREASSEELKAAVRLCMLYHPDKHRDPDLKSOAERLNF 60  
D 1 MATALSEELDNEDYYSLNVRREASSEELKAAVRLCMLYHPDKHRDPDLKSOAERLNF 60

```
QY 61 LVHQAEEVLSDPQTRAIYDIYKRGKLEMEGMEVVERRRTPAIREFEERLQREBERRLQ 120
Db 61 LVHQAEEVLSDPQTRAIYDIYKRGKLDVEGMEVVERKRTPAIREFEERLQREBERRLQ 120
QY 121 QRTNPKGTISVGVADATDLFDRYDEYEDVS---GSSFPQIEINKMHSOSIEAPLTATD 176
Db 121 QRTNPKGTISVGVADATDLFDRYDEYEDYDEISGGGGGGGLPHIEINKMHSOSIEAPLTATD 180
QY 177 TALISGSLSTONGGGSINFALRRVTSAGKMGELBFGAGDLGGPLFGKLFENLTPRCF 236
Db 181 TAVLSGSLSTHNGGGSINFALRRVTSAGKMGEBVEFGAGDTGPGPLFGKLFENLTSPRCF 240
QY 237 VTTNCAIQFSSRGIRPGLTTLVARNLDKNTVGYLQWRMGIQSAMNTSIYRDTKTSHTFTVA 296
Db 241 TTHAGCGLQFSSRGIRPGLTTLVARNLDKNTMGTGLQWRMGQSSAMNTSIYRDTKSSHTFTFA 300
QY 297 LQIGIPHSFALISYQHKFODDDQTRVKGSLKAGPFGTVVEYGAERKISRSISVGAAVSVG 356
Db 301 VOLGIPHTFIMMSYQYKFODDQTKIKGSYKSGFPGTVVEYGAETKISRSISVGAATVSVG 360
QY 357 VPGVSLKVLKLNASQTYFFPHILTDQLPSAMPVATVGPLVYYPAMHRLTIKPYLRQK 416
Db 361 VPGVSLKVLKLNASQTYFFPHILTDQLPSAVFATVGPLVYLAIQRLVIRPYRAQ 420
QY 417 EKELEKORBSAATDVLOKQOASAVRLMOESVRRITAEBSHGLIIVANMYGKFFVNDK 476
Db 421 EGELEKORBSASADVAKKQOAEAVRLMOESVRRITAEBSHGLIIVANMYGKFFVNDK 480
QY 477 SRSEKVKYIDVTVPLQCLVYKSKLITLTKASKAGLPQFYPCYGEENKLVLYQFQVGLH 536
Db 481 SRGHERARVIVTVPLQCLVYKSKLITLTKASKAGLPQFYPCYGEENKLVLYQFQVGLH 540
QY 537 QVWVLDSEALRIPKSHRIDTD 558
Db 541 QVLCGTLEALRIPKSHRIDND 562

RESULT 6
ID 08C124 PRELIMINARY; PRT; 327 AA.
AC 08C124;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-OCT-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical DnaJ N-terminal domain containing protein.
GN E030019A03RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RL Nature 420:563-573(2002).
DR EMBL: AK089983; BAC1027.1;
DR MGD; MG1:244386; E030019A03RIK.
DR InterPro: IPR001623; DnaJ_N.
DR Pfam: PF00226; DnaJ_1.
DR SMART: SM00271; DnaJ_1.
DR PROSITE: PS00636; DnaJ_1;
DR PROSITE: PS50076; DnaJ_2; 1.
KW Hypothetical protein.
SQ SEQUENCE 327 AA; 37190 MW; 7019E0838E2A315C CRC64;
```

Query Match 58.7%; Score 1677; DB 11; Length 327;  
Best Local Similarity 98.8%; Pred. No. 3.3e-121;  
Matches 323; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

```
QY 1 MATALSEEDLNEDDYSLNVRREASSEELKAAYRLCMLYHPDKHROPBELKSOAERLFN 60
Db 1 MATALSEEDLNEDDYSLNVRREASSEELKAAYRLCMLYHPDKHROPBELKSOAERLFN 60
QY 61 LVHQAEEVLSDPQTRAIYDIYKRGKLEMEGMEVVERRRTPAIREFEERLQREBERRLQ 120
Db 61 LVHQAEEVLSDPQTRAIYDIYKRGKLEMEGMEVVERKRTPAIREFEERLQREBERRLQ 120
QY 121 QRTNPKGTISVGVADATDLFDRYDEYEDVS---GSSFPQIEINKMHSOSIEAPLTATD 180
Db 121 QRTNPKGTISVGVADATDLFDRYDEYEDYDEISGGGGGGGLPHIEINKMHSOSIEAPLTATD 180
QY 181 SGLSSTONGGGSINFALRRVTSAGKMGELBFGAGDLGGPLFGKLFENLTPRCFVTN 240
Db 181 SGLSSTONGGGSINFALRRVTSAGKMGELBFGAGDLGGPLFGKLFENLTSPRCFVTN 240
QY 241 CALQFSSRGIRPGLTTLVARNLDKNTVGYLQWRMGIQSAMNTSIYRDTKTSHTFTVA 300
Db 241 CALQFSSRGIRPGLTTLVARNLDKNTVGYLQWRMGIQSAMNTSIYRDTKTSHTFTVA 300
QY 301 IPHSFALISYQHKFODDDQTRVKGSLK 327
Db 301 IPHSFALISYQHKFODDDQTRVKGSLK 327

RESULT 7
ID 07SZY3 PRELIMINARY; PRT; 330 AA.
AC 07SZY3;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE SI:ZK105A6.1.1 (Novel protein) (Fragment).
GN SI:ZK105A6.1.1.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Sehra H.;
RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL845422; CAE30412.1; -.
FT NON TER 330 330
SQ SEQUENCE 330 AA; 37381 MW; 0B5D33B3C0BF84 CRC64;
```

Query Match 50.8%; Score 1452; DB 13; Length 330;  
Best Local Similarity 80.9%; Pred. No. 8.1e-104;  
Matches 267; Conservative 37; Mismatches 22; Indels 4; Gaps 1;

```
QY 1 MATALSEEDLNEDDYSLNVRREASSEELKAAYRLCMLYHPDKHROPBELKSOAERLFN 60
Db 1 MATALSEEDLNEDDYSLNVRREASSEELKAAYRLCMLYHPDKHROPBELKSOAERLFN 60
QY 61 LVHQAEEVLSDPQTRAIYDIYKRGKLEMEGMEVVERRRTPAIREFEERLQREBERRLQ 120
Db 61 LVHQAEEVLSDPQTRAIYDIYKRGKLDVEGMEVVERKRTPAIREFEERLQREBERRLQ 120
QY 121 QRTNPKGTISVGVADATDLFDRYDEYEDVS---GSSFPQIEINKMHSOSIEAPLTATD 176
Db 121 QRTNPKGTISVGVADATDLFDRYDEYEDYDEISGGGGGGGLPHIEINKMHSOSIEAPLTATD 180
QY 177 TALISGSLSTONGGGSINFALRRVTSAGKMGELBFGAGDLGGPLFGKLFENLTSPRCF 236
Db 181 TAVLSGSLSTHNGGGSINFALRRVTSAGKMGEBVEFGAGDTGPGPLFGKLFENLTSPRCF 240
QY 237 VTTNCAIQFSSRGIRPGLTTLVARNLDKNTVGYLQWRMGIQSAMNTSIYRDTKTSHTFTVA 296
Db 241 TTHAGCGLQFSSRGIRPGLTTLVARNLDKNTMGTGLQWRMGQSSAMNTSIYRDTKSSHTFTFA 300
QY 297 LQIGIPHSFALISYQHKFODDDQTRVKGSL 326
Db 297 LQIGIPHSFALISYQHKFODDDQTRVKGSL 326
```



Db 301 VOLGIPHTFIMMSYQKFODDQTKIKSV 330

RESULT 8

Q9UM08 PRELIMINARY; PRT; 318 AA.

AC Q9UM08; (Tremblrel. 13. Created)

DT 01-MAY-2000 (Tremblrel. 13. Last sequence update)

DT 01-MAY-2000 (Tremblrel. 13. Last sequence update)

DT 01-MAR-2003 (Tremblrel. 23. Last annotation update)

DE Hypothetical protein.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RA Rhodes S., Huckle E.;

RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL; AL109978; CAB53376.1; -

DR HSSP; P25685; 1HDJ.

DR InterPro; IPR001623; DnaJ\_N.

DR Pfam; PF00226; DnaJ\_1.

DR SMART; SM00271; DnaJ\_1.

DR PROSITE; PS00636; DnaJ\_1; 1.

DR PROSITE; PS50076; DnaJ\_2; 1.

KM Hypothetical protein.

SC SEQUENCE 318 AA; 35610 MW; F287BD843EBB94 CRC64;

Query Match 42.7%; Score 1220; DB 4; Length 318;

Best Local Similarity 100.0%; Pred. No. 6,6e-66;

Matches 234; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 39 MLVHPDGRDELKSGAERLNLVHQAIVLSDPQTRAIYDIYKRGLEMGWEVERRR 98

DB 1 MLVHPDGRDELKSGAERLNLVHQAIVLSDPQTRAIYDIYKRGLEMGWEVERRR 60

QY 99 TPARIREFEFLQREERRRRLQORTNPKGTSVGVADTDLFDRDEDEVEDVSSGSPQIE 158

DB 61 TPARIREFEFLQREERRRRLQORTNPKGTSVGVADTDLFDRDEDEVEDVSSGSPQIE 120

QY 159 INKNIHQSIIEAPLTATDTALISGLSTONGGGSINFALRYTSKAGMGELEFGAGDL 218

DB 121 INKNIHQSIIEAPLTATDTALISGLSTONGGGSINFALRYTSKAGMGELEFGAGDL 180

QY 219 QGPIFGKLFNLTFRCTVTTCALQPSRGIRPGLTTVARNLDKNTVGLQW 272

DB 181 QGPIFGKLFNLTFRCTVTTCALQPSRGIRPGLTTVARNLDKNTVGLQW 234

RESULT 9

Q9V6Z8 PRELIMINARY; PRT; 545 AA.

AC Q9V6Z8; Q8T054;

DT 01-MAY-2000 (Tremblrel. 13. Created)

DT 01-MAY-2000 (Tremblrel. 22. Last sequence update)

DT 01-MAR-2003 (Tremblrel. 23. Last annotation update)

DE CG5531 protein (LUD27406p).

GN CG5531.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydriidae; Drosophilidae; Drosophila.

OX NCBI\_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RA STRAIN=Berkeley;

RX MEDLINE=20196005; PubMed=10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D., Amaratunga C.P., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F., George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N., Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X., Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,

Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G., Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D., Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M., Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolintsov S., Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotier P., Burks K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I., Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P., de Pablo B., Delcher A., Deng Z., May A.D., Dew I., Dietz S.M., Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P., Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischman W., Foster C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glasser K., Glodek A., Gong P., Gorrell J.H., Gu Z., Guan P., Harris M., Hartley L., Harvey D., Heiman T.J., Hernandez J.R., Houck J., Houston D., Houston K.A., Howland T.J., Wei M.-H., Iegem C., Jaitai M., Kalush F., Karpen G.H., Ke Z., Kenson J.A., Ketchum K.A., Kimmel B.B., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z., Lasco P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X., Liu X., Mattei B., McIntosh T.C., Melrod M.P., McPherson D., Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A., Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L., Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Paclet J.M., Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G., Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H., Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T., Spier E., Spradling A.C., Stapleton M., Strong R., Sun E., Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X., Wang Z.-Y., Wasmann D.A., Weinstein G.M., Weissbach J., Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L., Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O., Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

RT "the genome sequence of Drosophila melanogaster.";

RL Science 287:2185-2195(2000).

RN [2]

RP SEQUENCE FROM N.A.

RA Celniker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A., Evans C.A., Gocayne J.D., Amaratunga C.P., Brandon R.C., Rogers Y., Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A., Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M., Dodson K., Dorett V., Doup L.E., Doyle C., Dresnek D., Farfan D., Ferreira S., Frise E., Galle R.F., Gary N.S., George R.A., Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J., Iegem C., Jaitai M., Kruse D., Li P., Mattei B., Moshrefi A., McInosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J., Paclet J., Parag V., Park S., Patel S., Pfeiffer B., Pounanavong S., Pittman G.S., Puri V., Richards S., Scheeler F., Stapleton M., Strong R., Svirskas R., Tector C., Tyler D., Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;

RT "Sequencing of Drosophila melanogaster genome.";

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE FROM N.A.

RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K., Hradecky P., Huang Y., Kaminker J.S., Prochorko S.E., Smith C.D., Tuhy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E., Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N., Krommiller B., Marshall B., Millburn G., Richter J., Russo S., Seale S.M.J., Smith E., Shu S., Smutnak F., Whitfield E., Ashburner M., Gelbart W.M., Rubin G.M., Wengell C.J., Lewis S.E.;

RT "Annotation of Drosophila melanogaster genome.";

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RN [4]

RP SEQUENCE FROM N.A.

RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RN [5]

RP SEQUENCE FROM N.A.

RA FlyBase;

RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.

RN [6]

RP SEQUENCE FROM N.A.

RA STRAIN=Berkeley;

RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,

RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,  
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,  
RA Nuno J., Pacle J., Paragas V., Park S., Phouanavong S., Wan K.,  
RA Yu C., Lewis S.E., Rubin G.M., Celisner S.;  
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AEO03815; AAF58269.2; -;  
DR EMBL; AYO69547; AAL39692.1; -;  
DR HSSP; P25685; 1HDJ.  
DR FlyBase; FBgn0033918; CG8531.  
DR InterPro; IPR001623; DnaJ\_N.  
DR Pfam; PF002226; DnaJ; 1.  
DR SMART; SM00271; DnaJ; 1.  
DR PROSITE; PS50076; DnaJ\_2; 1.  
SQ SEQUENCE 545 AA; 60912 MW; F605485B094CDD06 CRC64;

Query Match 40.6%; Score 1160.5; DB 5; Length 545;  
Best Local Similarity 43.9%; Pred. No. 66-81; Indels 13; Gaps 6;  
Matches 239; Conservative 116; Mismatches 176;

QY 6 SEBELNEDYYSLLNVRREASSEELKAAYRLCMLYHPDKHDPBELKQAEFLNVLVHQ 65  
8 SDAELO-ENYTFELNPRDATABQINTAYRKQSMFHPDKHLPDSKMAEIMFNFTKA 66  
DB 66 YEVLSDPQRAIYDVGKLEMEGWEVERRRTPAIREEFRLQREBERLQORNP 125  
67 YEVLSDPQRAIYDVGKLEMEGWEVERRRTPAIREEFRLQREBERLQORNP 126  
QY 126 KGTISVGDATDLDPRDEEYDVSGSSPQIEINMKHSIQSIEAPLTATDTAILSGSIS 185  
127 KGTITIVNATEIPAPIDD-----SEMPHEIGSMISIQSIEAPLTRKDMIMSGNLY 179  
DB 186 TONGNGGSGINFALRVTSAGKGELEFGADLQGPLFKLPRLNTPRCFTVTTNCALO 245  
180 SSNGSGGSGFVIGRRILN-KGWEICAGAGN--GFLGKGGRTLSQKTLINGTNLNL 236  
QY 246 SSGIRGRLTTLARNIDKRTVGYLQWRNGIQSANMSTYRDTKTSHTFVALOLGIPHS 305  
237 RDGQVLPALFSTLAVOLDKHTMGSLLTNAGSOSMSFQIDHSKETYSLSSSLVIGTPHY 296  
QY 306 ALISYGHKFDODDQTVRKSGSLKAGFCTVVEYGEKRTKSHSVLGAIVSGVPOGSLKY 365  
297 FGLSYTRKMM-ENELKIKLAKAGTGFGEYGEKKSIVTSVITVSGVSVILKF 355  
QY 366 KLRASQTYFPPIHLDOLLPAMFYATVGPLVYVYFAMRLIKPYLRAQEKELKOR 425  
356 KILRSNGSYVEPIHLSDEIVPAVAVSVTPVIAWPFIKRTWDPMEAKNKIEVERTK 415  
QY 426 SAATDVLOKQKESAVRLMOESVRRIIEESRMGLIYNANYGKFAVNDKSKSEKVKY 485  
416 QNEQRISAKHEASAAVHLMQATYRNIMTEELARNGLITRAVYGGCTLAG-GRQFKPEOS 474  
QY 486 IDVTPLCLVKSQKILITLTSKAGLPGFPCVGEKKNLKVLYORFGLVHGVVLDSEA 545  
475 LDVTVAIQCVNKGKGTQLHDSKSDLPGFYDPIGDKILIRIETTYONQPEVHIHNDNA 534  
DB 546 LRIP 549  
535 LRIP 538

RESULT 10  
Q8BP83 PRELIMINARY; PRT; 214 AA.  
AC O8BP83;  
DT 01-MAR-2003 (TREMBLrel. 23, Created)  
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)  
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)  
DE Hypoetical DnaJ N-terminal domain containing protein  
DE (fragment).  
GN E030019A03RIK.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_Taxid=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Body;  
RX MEDLINE=22354683; PubMed=12466851;  
RA The FANTOM Consortium,  
RA The RIKEN Genome Exploration Research Group Phase I & II Team;  
RT "Analysis of the mouse transcriptome based on functional annotation of  
RT 60,770 full-length cDNAs";  
RL Nature 420:563-573 (2002).  
DR EMBL; AK077544; BAC36856.1; -;  
DR MGD; MGI:2443386; E030019A03RIK.  
KW Hypoetical protein.  
FT NON TER 1  
SQ SEQUENCE 214 AA; 24077 MW; 4D2C1A23D882D88 CRC64;

Query Match 35.8%; Score 1022; DB 11; Length 214;  
Best Local Similarity 93.5%; Pred. No. 7,2e-71;  
Matches 200; Conservative 11; Mismatches 3; Indels 0; Gaps 0;

QY 346 HSYLGAASVGPVGVSLSKLNLRASQTYFPPIHLDOLLPAMFYATVGPLVYFAMR 405  
1 HSYLGAASVGPVGVSLSKLNLRASQTYFPPIHLDOLLPAMFYATVGPLVYFAMR 60  
QY 406 LIKPYLRAQEKELKORSAATDVLOKQKESAVRLMOESVRRIIEESRMGLIIV 465  
61 LIKPYLRAQEKELKORSAATDVLOKQKESAVRLMOESVRRIIEESRMGLIIV 120  
QY 466 NAMYGKFNVDKSRKKEKXVIDVTVPLQCLVDSKTLITLTSKAGLPGFYDVCVEKNTL 525  
121 NAMYGKFNVDKSRKKEKXVIDVTVPLQCLVDSKTLITLTSKAGLPGFYDVCVEKNTL 180  
QY 526 KYLYOFRGVLHGVVLDSEALRIPKQSHRIDTDG 559  
181 KYLYOFRGVLHGVVLDSEALRIPKQSHRIDTDG 214  
DB

RESULT 11  
Q8LN96 PRELIMINARY; PRT; 569 AA.  
AC Q8LN96;  
DT 01-OCT-2002 (TREMBLrel. 22, Created)  
DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)  
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
DE Putative DnaJ domain containing protein.  
GN OSUNB0015K05.11.  
OS Oryza sativa (japonica cultivar-group).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC Ehrhartoideae; Oryzae; Oryza.  
OX NCBI\_Taxid=39947;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Nipponbare;  
RA Buell C.R., Yuan Q., Ouyang S., Liu J., Ganeberg K., Jones K.M.,  
RA Overton I.I., Tsitrin T., Kim M.M., Bera J.U., Jin S.S.,  
RA Fadrosch D.W., Tallon L.J., Koo H., Ziemann V., Heisio J., Blunt S.,  
RA Vanaken S.S., Riedmuller S.B., Uteback T.T., Feldlyum T.V.,  
RA Yang Q.Q., Haas B.J., Suh B.B., Peterson J.J., Quakebush J.,  
RA White O., Salzberg S.L., Fraser C.M.,  
RT "Oryza sativa chromosome 10 BAC OSUNB0015K05 genomic sequence.",  
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Nipponbare;  
RA The Rice Chromosome 10 Sequencing Consortium;  
RT "In-depth view of structure, activity, and evolution of rice  
RT chromosome 10.";  
RL Science 300:1566-1569 (2003).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Nipponbare;  
RA Buell C.R., Wing R.A., McComb W.R., Messing J., Yuan Q.;

RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.

DR EMBL; AC090870; AAM95682.1; -  
DR EMBL; AB017110; AAP54538.1; -  
DR Gramene; O8LN96; -  
DR InterPro; IPR001623; DnaJ\_N.  
DR Pfam; PF00226; DnaJ\_1.  
DR SMART; SM00271; DnaJ\_1.  
DR PROSITE; PS00636; DnaJ\_1; 1.  
DR PROSITE; PS50076; DnaJ\_2; 1.  
DR PROSITE; PS50076; DnaJ\_2; 1.  
SQ SEQUENCE 569 AA; 63322 MW; B5D44574DBE415AB CRC64;

Query Match 23.1%; Score 661; DB 10; Length 569;  
Best Local Similarity 30.2%; Pred. No. 2.6e-42;  
Matches 178; Conservative 103; Mismatches 242; Indels 66; Gaps 14;

QY 8 EELNEDYYSILNRRASSEELKAYRLCMLTHPDGRPELKSQARLFLNVHQA 67  
DB 6 EPEGRRELIALHSPDASGEIRRAYQVQIYHPDKYQDPOMKDVATENFORIRDA 65  
QY 68 VLSPDQTRAIYDIYGRGLMEGMEVEVERRRTPAIREFERLQREERRLQQRTPNK 127  
DB 66 ILSDENKQIYDIYGMELN-SGLELGPKNKPEIKQLEKRRKEEKFLLAHARTG 124  
QY 128 TISVGVADTDLFDYDEEYEDVSGSFPQIIBINKMHSQSIAPLTAITAILSGSLSTQ 187  
DB 125 SIANFVSPOYLDGYG-----IMRGKMSSEVQLPFSKKNVVGNNLVN 170  
QY 188 NGNGGGSINPALRYVTSKAGGELFPGAGDLOGPLFGKLFRNLTPRCFTVTTCALOPSS 247  
DB 171 GTDTGAASAVLRHQLSSV--ASVEFMATAGLRSLISVQTRFOISPHSTATSGALSLRD 228  
QY 248 RGIRPGLTTVLARNLDKRTVGVQLQWRMGIOGAMNTSIYDRKTSHTFVALQGIHPSPAL 307  
DB 229 GSI--NLNMTATRLQSDNIYVIGNIQLALGTDSSISVGMKQDEKKSAAQDVKLGNVFGAS 286  
QY 308 ISYHKFPDDQTRVYKSLKAGFPGTVVEYGAERKISRHSVILGAVSGV-----358  
DB 287 AHYRYFSTKSHGRAVG--RVGSTALDPEIGGRRISFSTVRMIVNIGINISILEDT 344  
QY 359 -----QGVSLKVKLNKASQTYFFPIHLT---DQLSPAMYATVGP 396  
DB 345 VINKGMIKWESIYIGEVEBGSWRFELHAGQKLVIPVLLSTDFNALATSPFAL--P 401  
QY 397 LVVYFAMHRLTIKPY-LPAQKEKELEKORESAATDVLOKKOASAVRLMESRRIIEA 455  
DB 402 STLFLLQTYFVVKCYCKREKQKELEKM-ESLSQLEAPRAKQAQCLLEPVSNNKKNR 460  
QY 456 EESRMGLIIVNMYG--KFVNDKSRKSE-----KVKVLDVTVPLQCLVQKS-KLILTEA- 506  
DB 461 QLEDDGLVITKALYGNRKVKESSESNELNDVASQVLDVTIIPLNFVSEAGQKLHGI 520  
QY 507 SKAQLPGFYDPCVGESEKMLKVLQYQRGVLAQVNVLDSEALRIPKQSHRI 555  
DB 521 KKSIGMGFYDPCGDPKLLVETFGQYKVMVDYALILPQIHQI 569

## RESULT 12

AC P91243 PRELIMINARY; PRT; 591 AA.  
DT 01-MAY-1997 (TREMblrel. 03; Created)  
DT 01-MAY-1997 (TREMblrel. 03; Last sequence update)  
DT 01-OCT-2003 (TREMblrel. 25; Last annotation update)  
DE Hypothetical protein.  
GN F11G11.7.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN 1;  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RX MEDLINE=99069613; PubMed=9851916;

RA None;  
RT "Genome sequence of the nematode C. elegans: a platform for  
RT investigating biology. The C. elegans Sequencing Consortium.";  
RL Science 282:2012-2018(1998).

RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA latreille P.; Deadman R.;  
RT "The sequence of C. elegans cosmid F11G11.";  
RT Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA Waterston R.;  
RT "Direct Submission.";  
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.

QY 7 EELNEDYYSILNRRASSEELKAYRLCMLTHPDGRPELKSQARLFLNVHQA 66  
DB 22 EEBI---DFYALINVPDADTDDEIRKAYRRCMLFDRVDNDEKDAERVFVKLRRAH 78  
QY 67 EVLSDPQTRAIYDIYGRGLMEGMEVEVERRRTPAIREFERLQREERRLQQRTPNK 126  
DB 79 EVLSDPQTRAIYDIYGRGLMEGMEVEVERRRTPAIREFERLQREERRLQQRTPNK 138  
QY 127 GTISGVADTDLFDYDEEYEDVSGSFPQIIBINKMHSQSIAPLTAITAILSGSLST 186  
DB 139 SAFMIKTTIGMFQENED-----RYRP-QLLGISLSGSDVCAPFGVDRPFGSGVKT 190  
QY 187 QNNGGGSINPALRYVTSKAGGELFPGAGDLOGPLFGKLFRNLTPRCFTVTTCALOPSS 247  
DB 191 GNGRGDSVSAVWKRVA-----GSYNLENTISLSASVSLTCRAARNVFTTRA 240  
QY 239 TNCALOPS--SRGIRPGLTTVLARNLDK-----NTVGY---LQWR---274  
DB 241 VQPOLQYNMLHEALIPSIAMSMNLRIKTVYSEKRLFTDSSIFCKIIPFYSMRLHTRWQGS 300  
QY 275 -----GIGAMNTSIYDRKTSHTFVALQI---GIHPSPALISYGHKFPDDQTRVYK 325  
DB 301 IVNLISPIANALTTTVH--TENNHAQVGLTSPINSNVLVYIIRKPNDSITTEM--S 357  
QY 326 LKAGFPGTVVEYGAERKISRHSVILGAVSGVQ--VSLKVKLNKASQTYFFPIHLT--382  
DB 358 VQLTGGVGNPAISMDRRLSRYSRISCSFHSFSCLLYTFKCLKAGGSTFDMQIVLCDDK 417  
QY 383 QLLPSAMFYATVGPVLVYFA--MHRLLIKPYLAQKEKELEKORESAATDVLOKKOAS 440  
DB 418 EALSRSVLY--GVALPYFSFOIAKYIFRMWKEKFSMEDNSREQVD--VAKKEBAAN 472  
QY 441 AVRIMQESVRIIEAESRMGLIIVNMYGKFVNDKSRKSEKV--KVLDVTVPLQCLVQ 498  
DB 473 IVSLMATARIRGDESKGVILISAKQGCVCVSGTRAVPLAGEHTIDVTVLQAMVD 532  
QY 499 SKLILTEASAGLPGFYDPCVGESEKMLKVLQYQRGVLAQVNVLDSEALRIPKQSHRI 558  
DB 533 SOLRVYTV--KSQLPGFYDPCGDPKLLVETFGQYKVMVDYALILPQIHQI 591





**This page blank (uspio)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 13, 2004, 08:43:46 ; Search time 12299 Seconds

(without alignments)  
11322.967 Million cell updates/sec

Title: US-09-787-678A-12

Perfect score: 3213

Sequence: 1 gaagagctgcgaagatgagc.....gaagcagaaaaa 3213

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 2167151695 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenBank1.\*  
1: gb\_ba.\*  
2: gb\_hcg.\*  
3: gb\_in.\*  
4: gb\_om.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sbs.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vi.\*  
15: gb\_da.\*  
16: em\_hum.\*  
17: em\_hum.\*  
18: em\_in.\*  
19: em\_mu.\*  
20: em\_om.\*  
21: em\_or.\*  
22: em\_ov.\*  
23: em\_pat.\*  
24: em\_ph.\*  
25: em\_pl.\*  
26: em\_ro.\*  
27: em\_sbs.\*  
28: em\_un.\*  
29: em\_vi.\*  
30: em\_hcg\_hum.\*  
31: em\_hcg\_inv.\*  
32: em\_hcg\_other.\*  
33: em\_hcg\_mus.\*  
34: em\_hcg\_pln.\*  
35: em\_hcg\_rod.\*  
36: em\_hcg\_mam.\*  
37: em\_hcg\_vtc.\*  
38: em\_ey.\*  
39: em\_hcgo\_hum.\*  
40: em\_hcgo\_mus.\*  
41: em\_hcgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3191.4	99.3	3230	9	AF306695 Homo sapi
2	3171.6	98.7	3199	6	AK877081 Sequence
3	3171.6	98.7	3199	6	BD156474 Primer fo
4	3171.6	98.7	3199	9	AK001599 Homo sapi
5	3157.4	98.1	3178	9	HSMB805121
6	1947	60.6	2064	9	AK025491
7	1934	60.2	2125	9	BC014145
8	1851.6	57.6	2454	6	AX748271 Sequence
9	1851.6	57.6	2454	9	AK093709 Homo sapi
10	1845.2	57.4	3012	9	AK095386
11	1517.2	47.2	112158	9	HS126A5
12	1510.8	47.0	101481	2	AC108123
13	1397.4	43.5	2052	6	AX834019
14	1397.4	43.5	2052	9	AK096217
15	1236.4	38.3	1258	9	AF052158
16	1230.2	38.3	1258	2	AC062024
17	1006.6	31.3	1014	9	AF306694
18	989.8	30.8	2774	5	BC046897
19	832.2	25.9	1767	6	AK339329
20	746.2	23.2	1357	9	HS126A512
21	684.2	21.3	706	6	AX868425
22	684.2	21.3	779	6	BD221093
23	575.6	17.9	477	9	HUMZD67D08
24	466.4	14.5	455	6	BD264940
25	455	14.2	455	6	BD264940
26	455	14.2	455	6	AK238343
27	455	14.2	455	6	AX366559
28	445.2	13.9	550	6	AX873477
29	445.2	13.9	550	6	BD153539
30	385	12.0	386	6	BD221537
31	373	11.6	373	6	AX885956
32	373	11.6	373	6	BD025566
33	331.4	10.3	438	11	G37511
34	306	9.5	821	6	AX554944
35	302.6	9.4	455	6	AX330980
36	300	9.3	300	6	BD219127
37	292.2	9.1	2000	3	AY069547
38	271	8.4	355	9	AF306696
39	200.8	6.2	1411	3	AK115029
40	170.4	5.3	201635	10	AL611931
41	166.2	5.2	171	11	G62036
42	162.8	5.1	237743	2	AC129831
43	139.2	4.3	35217	2	AC017131
44	139.2	4.3	165176	3	AC007588
45	139.2	4.3	180425	3	AC008229

ALIGNMENTS

RESULT 1  
LOCUS AF306695 3230 bp mRNA linear PRI 01-OCT-2002  
DEFINITION Homo sapiens clone R5' mRNA sequence.  
ACCESSION AF306695  
VERSION AF306695.1 GI:23393219  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (Bases 1 to 3230)  
AUTHORS Xu,H., Sha,Y., Foltz,L., Romero,R., Larsen,A., Mahoney,W. and  
Schueler,P.A.  
TITLE Identification and characterization of genes expressed in human

JOURNAL fetal erythroblasts  
REFERENCE 2 (bases 1 to 3230)  
AUTHORS Xu, H., Sha, Y., Foltz, L., Romero, R., Larsen, A., Mahoney, W. and Schueler, P. A.  
TITLE Direct Submission  
JOURNAL Submitted (18-SEP-2000) Chief Technology Office, Roche Diagnostics, 2929 7th Street, Suite 100, Berkeley, CA 94710, USA  
FEATURES Location/Qualifiers  
source 1.3230  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="R5"  
/cell\_type="erythroblasts"  
/dev\_stage="fetal"

ORIGIN

Query Match 99.3%; Score 3191.4; DB 9; Length 3230;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 3206; Conservative 0; Mismatches 6; Indels 1; Gaps 1;

QY 1 GAAAGTTGCGAAGATGCGACGCGCTTGAGCGAGAGAGCTGGACATGAAGACTATT 60  
DB 7 GAAAGTTGCGAAGATGCGACGCGCTTGAGCGAGAGAGCTGGACATGAAGACTATT 66  
QY 61 ACTGTTGCTGAAAGTGGCGAGGGAGGCGCTTCTGAGAGCTGAAAGCTGCTACCGGA 120  
DB 67 ACTGTTGCTGAAAGTGGCGAGGGAGGCGCTTCTGAGAGCTGAAAGCTGCTACCGGA 126  
QY 121 GGCTGTGATGCTCTACCATCCAGACAGACAGAGAGCCAGAGCTCAAGTCAACAGCGG 180  
DB 127 GGCTGTGATGCTCTACCATCCAGACAGAGAGCCAGAGCTCAAGTCAACAGCGG 186  
QY 181 AAGAGCTGTTAACTTGTTCACAGAGCTTATGAGTGTAGTGAATGAGTGAATGAGG 240  
DB 187 AAGAGCTGTTAACTTGTTCACAGAGCTTATGAGTGTAGTGAATGAGTGAATGAGG 246  
QY 241 CCATCTATGATATATATGAGAGAGAGAGCTGGAATGAGAGAGTGGAGGTTGGAAA 300  
DB 247 CCATCTATGATATATATGAGAGAGAGAGCTGGAATGAGAGAGTGGAGGTTGGAAA 306  
QY 301 GAGAGAGAACCCCTGCTGAAATTCAGAGAGATTTGAGCGGCTGACAGAGAGAGAGAG 360  
DB 307 GAGAGAGAACCCCTGCTGAAATTCAGAGAGATTTGAGCGGCTGACAGAGAGAGAGAG 366  
QY 361 AAGAGAGATTCAG 420  
DB 367 AAGAGAGATTCAG 426  
QY 421 CCGACCTTTTGAATGCGTATGATGAGAGATGAGAGATGAGTGGAGTGGAGTGGAGTGG 480  
DB 427 CCGACCTTTTGAATGCGTATGATGAGAGATGAGAGATGAGTGGAGTGGAGTGGAGTGG 486  
QY 481 AGATTGAATTAATAAATGACATATCCAGTCCATTGAGGACCCCTTGACAGCGACAG 540  
DB 487 AGATTGAATTAATAAATGACATATCCAGTCCATTGAGGACCCCTTGACAGCGACAG 546  
QY 541 ACACAGCATTCCTCTCTGGAAGCTCTCAACCCAGAAATGGAATGGAAGGTTCCATT 600  
DB 547 ACACAGCATTCCTCTCTGGAAGCTCTCAACCCAGAAATGGAATGGAAGGTTCCATT 606  
QY 601 ACTTTGGCTCAGACGAGTAACCTTGGCAAAAGGAGTGGGAGATTTGGAAATTTGGAGCTG 660  
DB 607 ACTTTGGCTCAGACGAGTAACCTTGGCAAAAGGAGTGGGAGATTTGGAAATTTGGAGCTG 666  
QY 661 GAGACCTACAGAGGCGCTTTGTCGCTCAAGCTGTTCCGTAAATTCACACCAAGATGCT 720  
DB 667 GAGACCTACAGAGGCGCTTTGTCGCTCAAGCTGTTCCGTAAATTCACACCAAGATGCT 726  
QY 721 TTGTGACAAACAACTGTGCTCTGCAATTTTCATCCGTTGAATCCGAGCTGAGCA 780  
DB 727 TTGTGACAAACAACTGTGCTCTGCAATTTTCATCCGTTGAATCCGAGCTGAGCA 786

QY 781 CTGTCTAGCTGGAACTTAGCAAGAACACCGTGGGCTACCTGACAGTGGGCGATG 840  
DB 787 CTGTCTAGCTGGAACTTAGCAAGAACACCGTGGGCTACCTGACAGTGGGCGATG 846  
QY 841 TCCAGTACAGCAATGAACACTAGCATGTCGAGACATTAACCAAGCACTTCACTGTG 900  
DB 847 TCCAGTACAGCAATGAACACTAGCATGTCGAGACATTAACCAAGCACTTCACTGTG 906  
QY 901 CCTGAGCTGGAAATCCCTCACTCCTTTGACATGATCAGTATCAGCAAAATTCAG 960  
DB 907 CCTGAGCTGGAAATCCCTCACTCCTTTGACATGATCAGTATCAGCAAAATTCAG 966  
QY 961 ATGACATCAGACTCGTGTGAAGAGATCCCTCAAGCAGGCTTCTTTGGAGCGTGTG 1020  
DB 967 ATGACATCAGACTCGTGTGAAGAGATCCCTCAAGCAGGCTTCTTTGGAGCGTGTG 1026  
QY 1021 AGTACGAGCTGAGAGAAAGATCTCAGGACAGCGTTTGGGTGACGCTGCAAGCTTG 1080  
DB 1027 AGTACGAGCTGAGAGAAAGATCTCAGGACAGCGTTTGGGTGACGCTGCAAGCTTG 1086  
QY 1081 GAGTTCACAGGCGCTTCTTCAAAAGTCAAGTCAAGGCGCACTCAATCTTCT 1140  
DB 1087 GAGTTCACAGGCGCTTCTTCAAAAGTCAAGTCAAGGCGCACTCAATCTTCT 1146  
QY 1141 TCCCTATTCACTTGAACGACAGCTTCTGCGAGCGCATGTTCTATGSCACCGTGGG 1200  
DB 1147 TCCCTATTCACTTGAACGACAGCTTCTGCGAGCGCATGTTCTATGSCACCGTGGG 1206  
QY 1201 CTCTAGTGTCTACTTTGCGATGACCGCTGATCATCAAACTCACTCAGGCGTCA 1260  
DB 1207 CTCTAGTGTCTACTTTGCGATGACCGCTGATCATCAAACTCACTCAGGCGTCA 1266  
QY 1261 AAGAGAGAAATTTGAG 1320  
DB 1267 AAGAGAGAAATTTGAG 1326  
QY 1321 AAGAGCGAGAGTCCGCTGTCCGCTGATGACAGAAATCTGTCCGAAGATTAATTGAG 1380  
DB 1327 AAGAGCGAGAGTCCGCTGTCCGCTGATGACAGAAATCTGTCCGAAGATTAATTGAG 1386  
QY 1381 AAGAGTCCAGAAATGGGCTCATCATGTCATATGCTGTGACGGAAGTTGTCAATGAC 1440  
DB 1387 AAGAGTCCAGAAATGGGCTCATCATGTCATATGCTGTGACGGAAGTTGTCAATGAC 1446  
QY 1441 AAGAGAGAAAGCGAG 1500  
DB 1447 AAGAGAGAAAGCGAG 1506  
QY 1501 TGAAGGACTCGAAGCTCATCTCAAGAGGCTCCAAAGGCTGGGCTGCTGCTTTATG 1560  
DB 1507 TGAAGGACTCGAAGCTCATCTCAAGAGGCTCCAAAGGCTGGGCTGCTGCTTTATG 1566  
QY 1561 ACCCGTGTGGGGGAGAGAGAAACCTGAAAGTGTCTATCAGTTCCGGGGCGTCTGC 1620  
DB 1567 ACCCGTGTGGGGGAGAGAGAAACCTGAAAGTGTCTATCAGTTCCGGGGCGTCTGC 1626  
QY 1621 ATCAGGTATGCTGTGACATGAGGCGCTTCCGATACCAAGAGTGTCCACAGATG 1680  
DB 1627 ATCAGGTATGCTGTGACATGAGGCGCTTCCGATACCAAGAGTGTCCACAGATG 1686  
QY 1681 ATACAGATGATTAACCTGCAAGAACCAATTTTAAAGGCGCAAAATCTTTCT 1740  
DB 1687 ATACAGATGATTAACCTGCAAGAACCAATTTTAAAGGCGCAAAATCTTTCT 1746  
QY 1741 GGGAGTCTAACAAATTTGAAATGAAGAAACCCAGACATCAGATGTTTATTTATATTA 1800  
DB 1747 GGGAGTCTAACAAATTTGAAATGAAGAAACCCAGACATCAGATGTTTATTTATATTA 1806  
QY 1801 TTATTTAAGAGTGTACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1860  
DB 1807 TTATTTAAGAGTGTACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1866



QY 1861 TTGAGGGTGTGGGGGTAGACTGAGGACGCCCACTGGAGAACGAGCTGAGGCTGGCC 1920  
 Db 1867 TTGAGGGTGTGGGGGTAGACTGAGGACGCCCACTGGAGAACGAGCTGAGGCTGGCC 1926  
 QY 1921 CAGGCTGTGTTTCCCAAGATCACTTCTGGAGGAGGAGGCTGAGGCTGAGCTGGCTG 1980  
 Db 1927 CAGGCTGTGTTTCCCAAGATCACTTCTGGAGGAGGAGGCTGAGGCTGAGCTGGCTG 1986  
 QY 1981 TGTCCCGAGCAACAGTGTGACCGGAGCGCCCGCTGTAGTTCTTGGCTGGCTGGA 2040  
 Db 1987 TGTCCCGAGCAACAGTGTGACCGGAGCGCCCGCTGTAGTTCTTGGCTGGCTGGA 2046  
 QY 2041 GGTGTCTGTGAGAGACCTGCGCTCACAAGAGAGCGTGAAGCACTTGTGAGTCAAGC 2100  
 Db 2047 GGTGTCTGTGAGAGACCTGCGCTCACAAGAGAGCGTGAAGCACTTGTGAGTCAAGC 2106  
 QY 2101 TGAACATGGGAAACAACCTGAAAAGCAGGAGGCTCCCGGTCAAGGAGAGCTGTGTG 2160  
 Db 2107 TGAACATGGGAAACAACCTGAAAAGCAGGAGGCTCCCGGTCAAGGAGAGCTGTGTG 2166  
 QY 2161 CTGGCTTCCCATGACCACTCTCTCTGTGAATATTAATGTCTGAATCTGAGAGATT 2220  
 Db 2167 CTGGCTTCCCATGACCACTCTCTCTGTGAATATTAATGTCTGAATCTGAGAGATT 2226  
 QY 2221 GCGGGTTTATATAAATGCTTTTATCTGAGAAACAAGGGTTTGGAAATTAGTGTCTT 2280  
 Db 2227 GCGGGTTTATATAAATGCTTTTATCTGAGAAACAAGGGTTTGGAAATTAGTGTCTT 2286  
 QY 2281 TTTTCCCATCTCCAGAGCTGTCTCAAGTATTCACCGGAGCCCTCGGCTTGGAGCAAGGT 2340  
 Db 2287 TTTTCCCATCTCCAGAGCTGTCTCAAGTATTCACCGGAGCCCTCGGCTTGGAGCAAGGT 2346  
 QY 2341 AGTGTAACTCCCGATCCAGAGGCTTGAAGCTTGAACAGAGTGGCTTCCGATTCGGGTGG 2400  
 Db 2347 AGTGTAACTCCCGATCCAGAGGCTTGAAGCTTGAACAGAGTGGCTTCCGATTCGGGTGG 2406  
 QY 2401 GAAAACGCGCTGCGCAACGAGGAGCTGAGTGGCGGTGTGCTCCCTCAACCGGCTGACAC 2460  
 Db 2407 GAAAACGCGCTGCGCAACGAGGAGCTGAGTGGCGGTGTGCTCCCTCAACCGGCTGACAC 2466  
 QY 2461 CCACTCCAGAGTGTGAGTGTGGGCAAGGAGCTCAAGAGAGACAGACAGGCGCTTGG 2520  
 Db 2467 CCACTCCAGAGTGTGAGTGTGGGCAAGGAGCTCAAGAGAGACAGACAGGCGCTTGG 2526  
 QY 2521 CAAGCATCAGACACACCAACCAAGGCTGGAACCCAGAGCGCGCTGTGTATCCCA 2580  
 Db 2527 CAAGCATCAGACACACCAACCAAGGCTGGAACCCAGAGCGCGCTGTGTATCCCA 2586  
 QY 2581 GCAGGTGCACTGAGAGTGTGCGCTCTGCAAGTGTCAAGCGTCTCAAGAGAACACCAAGG 2640  
 Db 2587 GCAGGTGCACTGAGAGTGTGCGCTCTGCAAGTGTCAAGCGTCTCAAGAGAACACCAAGG 2646  
 QY 2641 CCTGTGCTCCGAGAGCTTCTTTCAGAACCTTCTCTCAGTGTCCCACTTGGGATCAGAT 2700  
 Db 2647 CCTGTGCTCCGAGAGCTTCTTTCAGAACCTTCTCTCAGTGTCCCACTTGGGATCAGAT 2706  
 QY 2701 GCAGCGAGTGAAGACCCCTCTCAAGCGCTGTGAACCTCGGCTGCAATTAAGTTCGTAGG 2760  
 Db 2707 GCAGCGAGTGAAGACCCCTCTCAAGCGCTGTGAACCTCGGCTGCAATTAAGTTCGTAGG 2766  
 QY 2761 CCTGTCTGTGGGGGCTGGAAGTGGAGCACTAGTGTCTTGTGTGACCCCTGGAGCA 2820  
 Db 2767 CCTGTCTGTGGGGGCTGGAAGTGGAGCACTAGTGTCTTGTGTGACCCCTGGAGCA 2826  
 QY 2821 AGCGCGGACAGTGTGTGCTGAAGCAGCTGCGCGGCGGAGGCCCAAGCTGCGCGGCT 2880  
 Db 2827 AGCGCGGACAGTGTGTGCTGAAGCAGCTGCGCGGCGGAGGCCCAAGCTGCGCGGCT 2886  
 QY 2881 CCAAGCCCAACCAAGCTTGTGTGAAGTCAAGCCCACTCCCAAGCACTGTATCTGAGT 2940  
 Db 2887 CCAAGCCCAACCAAGCTTGTGTGAAGTCAAGCCCACTCCCAAGCACTGTATCTGAGT 2946  
 QY 2941 AAGGCTTAAGAACCTTCTGTGTGTTTGAAGAGAGTTCGGGTGTCCAATTCTGTA 3000

Db 2947 AAGGCTTAAGAACCTTCTGTGTGTTGAAGAGAGTTCGGGTGTCCAATTCTGTA 3006  
 QY 3001 ACATTATCTTCATTTTAAAAAGGTTTCTGTACAGGCGCCCAAGCGCCGAGCGGGT 3060  
 Db 3007 ACATTATCTTCATTTTAAAAAGGTTTCTGTACAGGCGCCCAAGCGCCGAGCGGGT 3065  
 QY 3061 GAGCGGTGTGTGATGAGCTGGGCGCGGGGCTTCCCGTGGCGCTGTGCGAGAGTCT 3120  
 Db 3066 GAGCGGTGTGTGATGAGCTGGGCGCGGGGCTTCCCGTGGCGCTGTGCGAGAGTCT 3125  
 QY 3121 TCTGGGACCCCATCTTGTGCGTTTCATTTGCACTGACGACTGACAGAAAGGACTACACCA 3180  
 Db 3126 TCTGGGACCCCATCTTGTGCGTTTCATTTGCACTGACGACTGACAGAAAGGACTACACCA 3185  
 QY 3181 ATAAACCTTCTGAAAGCAGAAAAA 3213  
 Db 3186 ATAAACCTTCTGAAAGCAGAAAAA 3218  
 RESULT 2  
 AX877081 3199 bp DNA linear PAT 17-DEC-2003  
 LOCUS AX877081  
 DEFINITION Sequence 11986 from Patent EP1074617.  
 ACCESSION AX877081  
 VERSION AX877081.1 GI:40031817  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Homo sapiens (human)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE  
 1  
 AUTHORS Oca, T., Isogai, T., Nishikawa, T., Hayashi, K., Saito, K., Yamamoto, J.,  
 Ishii, S., Sugiyama, T., Wakamatsu, A., Nagai, K. and Otsuki, T.  
 TITLE Primers for synthesizing full-length cDNA and their use  
 JOURNAL Patent: EP 1074617-A 11986 07-FEB-2001;  
 Research Association for Biotechnology (JP)  
 FEATURES  
 source  
 1..3199  
 /organism="Homo sapiens"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:9606"  
 16..1695  
 /note="unnamed protein product"  
 /codon\_start=1  
 /protein\_id="CA889562.1"  
 /db\_xref="GI:40031818"  
 /translation="MATALSSEELDNDYSLNVRREASSEELKAYRRLCMLYHPD  
 KHRDPEKSGAERLAFNLVHOAYEVLSDPOTRAIYDIYGRGLMEKGEVYRRTTPE  
 IREPERLQREBERRLQQRINKGTTISGVADTDLPDRDEYEDYSGSPQIEN  
 KMHISQIEAPLTAIDTALISGLSTONGGGSINFALRVTSAKMGSLERFAGDL  
 QGPFGKLPRLNLPFCFVTTCALPSSGIRPGTLVLAARMLDKNTMGLYLRMGCI  
 OSAMNTEIVEDTTSHTFVALQGIPIHSLIYOHKFDODDQTRVSGSLKAGFETV  
 VEYAEKERKISRHSVLGAASVGVPOGSLVKLNRASQTFPPLHLDOLPAMFVA  
 TVQPIVIVFAMHRLIITKPYLRAGEKELESEKREKRVADIVTYPLOGLYDSGLILTEASK  
 RIIBAESRNGULITVNAHYKFPNDKSRSEKREKRVADIVTYPLOGLYDSGLILTEASK  
 AGLPFDPVCEGKMLKVLVYFRGVLMVWLDSRLRIPKSHRIDTG"  
 ORIGIN  
 Query Match 98.7%; Score 3171.6; DB 6; Length 3199;  
 Best Local Similarity 99.7%; Pred. No. 0;  
 Matches 3188; Conservative 0; Mismatches 9; Indels 1; Gaps 1;  
 QY 2 AAAGTTGCGAAGTGCAGCGGCTTGAACGAGAGAGCTGCAATGAAGAATACTTA 61  
 Db 3 AAAGTTGCGAAGTGCAGCGGCTTGAACGAGAGAGCTGCAATGAAGAATACTTA 62  
 QY 62 CTGCTTCTGAAGCTGTGGGAGGAGGCTTCTTGAAGAGCTGAAGAGCTGACCGGAG 121  
 Db 63 CTGCTTCTGAAGCTGTGGGAGGAGGCTTCTTGAAGAGCTGAAGAGCTGACCGGAG 122  
 QY 122 GCTCTATGCTCTACCATCAAGACAGACAGAGACCCAGAGAGCTCAAGTCAAGCGGAG 181

Db 123 GCTCTGATGCTCTACCATCCAGCAAGCAAGACCCAGAGCTCAAGTCCAGGGCGA 182  
Qy 182 ACGACGTGTTAACTTGTTCCAGAGGCTTATGAAAGTCTTAAGTGAACCCCAAAACAGAGGC 241  
Db 183 ACGACGTGTTAACTTGTTCCAGAGGCTTATGAAAGTCTTAAGTGAACCCCAAAACAGAGGC 242  
Qy 242 CATCTATGATATATATGGAAGAGAGAGCTGGAATGGAAGATGGAGAGTTGTGGAAG 301  
Db 243 CATCTATGATATATATGGAAGAGAGAGCTGGAATGGAAGATGGAGAGTTGTGGAAG 302  
Qy 302 GAGAGAAACCCCTGCTGAAATTTGAGAGAGTTTGAAGGCTGAGAGAGAGAGAGA 361  
Db 303 GAGAGAAACCCCTGCTGAAATTTGAGAGAGTTTGAAGGCTGAGAGAGAGAGAGA 362  
Qy 362 GAGAGATGTCAGAGCGAAACCAATCCCAAGGGAGATCAGAGTTGAGATGATGTCAC 421  
Db 363 GAGAGATGTCAGAGCGAAACCAATCCCAAGGGAGATCAGAGTTGAGATGATGTCAC 422  
Qy 422 CGACCTTTTGTATCGCTATGATGAGAGATGGAAGATGTGTCCGGCAGTAGCTTTCCGA 481  
Db 423 CGACCTTTTGTATCGCTATGATGAGAGATGGAAGATGTGTCCGGCAGTAGCTTTCCGA 482  
Qy 482 GATTGAAATTAATAAATGCAATATCCAGTCCATTGAGGCAACCTTGAACAGCGACAGA 541  
Db 483 GATTGAAATTAATAAATGCAATATCCAGTCCATTGAGGCAACCTTGAACAGCGACAGA 542  
Qy 542 CACAGGCATCTCTCTGGAAGCCTCTCAACCGAAATGGAATGAGAGAGTTCCATTAA 601  
Db 543 CACAGGCATCTCTCTGGAAGCCTCTCAACCGAAATGGAATGAGAGAGTTCCATTAA 602  
Qy 602 CTTTGCCTCAGACGAGTAATCTCGGCAAGAGATGGGAGAGTTGGAAATTTGAGAGTGG 661  
Db 603 CTTTGCCTCAGACGAGTAATCTCGGCAAGAGATGGGAGAGTTGGAAATTTGAGAGTGG 662  
Qy 662 AGACCTACAGAGGGCTTTGTCCGTCTCAAGCTGTTCCGTAATCTCAACCAAGATGCTT 721  
Db 663 AGACCTACAGAGGGCTTTGTCCGTCTCAAGCTGTTCCGTAATCTCAACCAAGATGCTT 722  
Qy 722 TGTGACAAACAACCTGTGCTCGAGTTTTCATCCCGTGAATCCGAGCCGGCCCTGACAC 781  
Db 723 TGTGACAAACAACCTGTGCTCGAGTTTTCATCCCGTGAATCCGAGCCGGCCCTGACAC 782  
Qy 782 TGTCTAGCTCGAACCCTAGACAAGAACACCGTGGGCTACCTGCAAGTGGGATG 841  
Db 783 TGTCTAGCTCGAACCCTAGACAAGAACACCGTGGGCTACCTGCAAGTGGGATG 842  
Qy 842 CCAAGTCAAGCACTAGCATCTGTCGAGACACTAAACCAAGCACTTCACTGTGGC 901  
Db 843 CCAAGTCAAGCACTAGCATCTGTCGAGACACTAAACCAAGCACTTCACTGTGGC 902  
Qy 902 CCTGCAGCTGGGAATCCCTCACTCTTGTGACTGATCAGCTATCAGCAAAATTTCCAGA 961  
Db 903 CCTGCAGCTGGGAATCCCTCACTCTTGTGACTGATCAGCTATCAGCAAAATTTCCAGA 962  
Qy 962 TGAAGATCAGACTCTGTGAAAGGATCCCTCAAGCAGGCTTCTTTGGGACGGTGTGGA 1021  
Db 963 TGAAGATCAGACTCTGTGAAAGGATCCCTCAAGCAGGCTTCTTTGGGACGGTGTGGA 1022  
Qy 1022 GTACGAGCTGAGAGGAATCTTCAGGCAAGCGTTTGGGTGACGCTGTCAAGCTTGG 1081  
Db 1023 GTACGAGCTGAGAGGAATCTTCAGGCAAGCGTTTGGGTGACGCTGTCAAGCTTGG 1082  
Qy 1082 AGTTTCAAGAGGGGTTTCTCTCAAAAGTCAAGCTCAACAGGGCCAGTCAATCTTCTT 1141  
Db 1083 AGTTTCAAGAGGGGTTTCTCTCAAAAGTCAAGCTCAACAGGGCCAGTCAATCTTCTT 1142  
Qy 1142 CCGTATTCATTGACGAGCAGCTTCTGCCAGCGCATGTTCTATGSCCAACCGTGGAGCC 1201  
Db 1143 CCGTATTCATTGACGAGCAGCTTCTGCCAGCGCATGTTCTATGSCCAACCGTGGAGCC 1202  
Qy 1202 TCTAGTGTCTTACTTTGSCATGCAACGCTGTGATCATCAACATCTCTCAGGCTCGAAG 1261  
Db 1203 TCTAGTGTCTTACTTTGSCATGCAACGCTGTGATCATCAACATCTCTCAGGCTCGAAG 1262

Qy 1262 AGAAGAGAAATTGAGAGACAGAGGGAAGCGCCGCCACCGATGTGCTGACAGAGAGCA 1321  
Db 1263 AGAAGAGAAATTGAGAGACAGAGGGAAGCGCCGCCACCGATGTGCTGACAGAGAGCA 1322  
Qy 1322 AGAGCGAGTCCGCTGTCGGCTGATGAGAGAAATCTGTCCGAGAGTAATTTGAGGAGA 1381  
Db 1323 AGAGCGAGTCCGCTGTCGGCTGATGAGAGAAATCTGTCCGAGAGTAATTTGAGGAGA 1382  
Qy 1382 AGAGTCCAGAAATGGGCTCATATGCTGCAATCCCTGTGTCGGGAAATTTGTCAATGACA 1441  
Db 1383 AGAGTCCAGAAATGGGCTCATATGCTGCAATCCCTGTGTCGGGAAATTTGTCAATGACA 1442  
Qy 1442 GAGCAGGAAGACGAGAAAGTGAAGTGAATGACGTAACGTAACCTGTGCCCTGACGTCTGTG 1501  
Db 1443 GAGCAGGAAGACGAGAAAGTGAAGTGAATGACGTAACGTAACCTGTGCCCTGACGTCTGTG 1502  
Qy 1502 GAAAGCTCGAAGCTCATCTCAACGAGGCTTCAAGGCTGGGCTGCTGCTTTTATGA 1561  
Db 1503 GAAAGCTCGAAGCTCATCTCAACGAGGCTTCAAGGCTGGGCTGCTGCTTTTATGA 1562  
Qy 1562 CCGGTGTGGGGGAAAGAGAACTGAAATGTCTTATCAAGTTCCGGGGGCTCTGCA 1621  
Db 1563 CCGGTGTGGGGGAAAGAGAACTGAAATGTCTTATCAAGTTCCGGGGGCTCTGCA 1622  
Qy 1622 TCAGGTGATGTGCTGGAAGTGAAGGCTCCCGGATACCAAGACAGTCCACAGATGCA 1681  
Db 1623 TCAGGTGATGTGCTGGAAGTGAAGGCTCCCGGATACCAAGACAGTCCACAGATGCA 1682  
Qy 1682 TACAGATGATTAACCTGCCAAGAACAGATTTTAAAGCCGCCAAAATCTTTTCTCG 1741  
Db 1683 TACAGATGATTAACCTGCCAAGAACAGATTTTAAAGCCGCCAAAATCTTTTCTCG 1742  
Qy 1742 GGAATCTAATAATTGGAAATGAAAAACCCGACATCGAGTGTTTTATTTATATAT 1801  
Db 1743 GGAATCTAATAATTGGAAATGAAAAACCCGACATCGAGTGTTTTATTTATATAT 1802  
Qy 1802 TATTATAGAAAGTGTGATACCATTAATCAATATGTAAGGAGCATGCAAGCAGCTT 1861  
Db 1803 TATTATAGAAAGTGTGATACCATTAATCAATATGTAAGGAGCATGCAAGCAGCTT 1862  
Qy 1862 TGAAGGTGCTGGGGTGAAGACTGAGGAGCCCACTGGGAACCAAGCTCAGCTGGCCC 1921  
Db 1863 TGAAGGTGCTGGGGTGAAGACTGAGGAGCCCACTGGGAACCAAGCTCAGCTGGCCC 1922  
Qy 1922 ATGCTGTTTTTCCCAAGATCAATTTCTGTGAGAGAAAGGCTCTGACCTGCTGCT 1981  
Db 1923 ATGCTGTTTTTCCCAAGATCAATTTCTGTGAGAGAAAGGCTCTGACCTGCTGCT 1982  
Qy 1982 GTCCGAGACAGTGTGACGAGCCCGCGCTGTGATGTTCTTGGCTGGGTCTGAG 2041  
Db 1983 GTCCGAGACAGTGTGACGAGCCCGCGCTGTGATGTTCTTGGCTGGGTCTGAG 2042  
Qy 2042 GTGTCTGTGAGACCTTGCTCTTCAACAGAGAGCTGAGCACTTCTGACGTCAAGCT 2101  
Db 2043 GTGTCTGTGAGACCTTGCTCTTCAACAGAGAGCTGAGCACTTCTGACGTCAAGCT 2102  
Qy 2102 GAAATATGGAACAACCTGTAAGAGAGAGGCTTCCGGTCAAGGAGGCTCTGCTGTGC 2161  
Db 2103 GAAATATGGAACAACCTGTAAGAGAGAGGCTTCCGGTCAAGGAGGCTCTGCTGTGC 2162  
Qy 2162 TGGCTTCCCATGACCACTCTCTGCTGTAATTAATGCTTGAATCTGAGACAGATG 2221  
Db 2163 TGGCTTCCCATGACCACTCTCTGCTGTAATTAATGCTTGAATCTGAGACAGATG 2222  
Qy 2222 CCGGTTTATTAACCTGCTTTTATCTGAGAACAAACGGGTGTAATTAATGCTGCTTT 2281  
Db 2223 CCGGTTTATTAACCTGCTTTTATCTGAGAACAAACGGGTGTAATTAATGCTGCTTT 2282  
Qy 2282 TTCCCACTCCAGAGCTGCTCAAGTCAATTCACGGGCTTCCGCTTGGGACAGGGTA 2341  
Db 2283 TTCCCACTCCAGAGCTGCTCAAAATCAATTCACGGGCTTCCGCTTGGGACAGGGTA 2342



Db 603 CTTGCGCTCAGACGAATACTTCGGCAAAAGGATGGGAGAGATTGGAAATTTGGAGCTGG 662  
QY 662 AGACCTACAGAGGGCTTTTGTGGTCTCAAGCTGTTCCGTAATCTCACACCAAGAGTCTT 721  
Db 663 AGACCTACAGAGGGCTTTTGTGGTCTCAAGCTGTTCCGTAATCTCACACCAAGAGTCTT 722  
QY 722 TGTGACAAACAACCTGTGCTCTGACATTTTTCATCCCGTGAATCCGACCCGGCTGACAC 781  
Db 723 TGTGACAAACAACCTGTGCTCTGACATTTTTCATCCCGTGAATCCGACCCGGCTGACAC 782  
QY 782 TGTCTAGCTCGGAACTTACACAAACACCCGTTGCTTCTGACATGTCGATGGGAT 841  
Db 783 TGTCTAGCTCGGAACTTACACAAACACCCGTTGCTTCTGACATGTCGATGGGAT 842  
QY 842 CCACTGACGCTGAAACCTGACATCGTCCGAGACCTTAAACAGGACCTTCACTGAGC 901  
Db 843 CCACTGACGCTGAAACCTGACATCGTCCGAGACCTTAAACAGGACCTTCACTGAGC 902  
QY 902 CCTGACAGCTGGAAATCCCTCACTCTTTTGACATGATCACTATCAGACCAAAATTTCCAGA 961  
Db 903 CCTGACAGCTGGAAATCCCTCACTCTTTTGACATGATCACTATCAGACCAAAATTTCCAGA 962  
QY 962 TGAAGATCAGATCTGTGTGAAGAGATCTCTCAAAAGCAGGCTTTCTTTGGAGCGGTGTGA 1021  
Db 963 TGAAGATCAGATCTGTGTGAAGAGATCTCTCAAAAGCAGGCTTTCTTTGGAGCGGTGTGA 1022  
QY 1022 GTACGGAGCTGAGAGGAATCTCCAGGCAACGGTTTTGGTGGACGCTGTCAAGCTGTG 1081  
Db 1023 GTACGGAGCTGAGAGGAATCTCCAGGCAACGGTTTTGGTGGACGCTGTCAAGCTGTG 1082  
QY 1082 AGTTCCACAGAGGGCTTTCTCTCAAAAGTCAAGCTCAACAGGGCCAGTCAGACATCTCTT 1141  
Db 1083 AGTTCCACAGAGGGCTTTCTCTCAAAAGTCAAGCTCAACAGGGCCAGTCAGACATCTCTT 1142  
QY 1142 CCTTATTCATTTGACGAGCAAGCTTCTGCCAGCGCATGTTCTATGCAACCGTGGGACC 1201  
Db 1143 CCTTATTCATTTGACGAGCAAGCTTCTGCCAGCGCATGTTCTATGCAACCGTGGGACC 1202  
QY 1202 TCTAGAGGTCTACTTTGGCATGACCGCTGATCATCAACCACTCTGAGGCTGAGAA 1261  
Db 1203 TCTAGAGGTCTACTTTGGCATGACCGCTGATCATCAACCACTCTGAGGCTGAGAA 1262  
QY 1262 AGAAGAGAAATTGGAG 1321  
Db 1263 AGAAGAGAAATTGGAG 1322  
QY 1322 AGAGGCGGAGTCCGCTGTCCGGCTGATGACGAGAACTGTCTCCAGAGATAATTGAGCAGA 1381  
Db 1323 AGAGGCGGAGTCCGCTGTCCGGCTGATGACGAGAACTGTCTCCAGAGATAATTGAGCAGA 1382  
QY 1382 AGAGTCCAGAAATGGGCTCTCATCGTCAATGCTGGTACGGGAAGTTGTCAATGCA 1441  
Db 1383 AGAGTCCAGAAATGGGCTCTCATCGTCAATGCTGGTACGGGAAGTTGTCAATGCA 1442  
QY 1442 GAGCAG 1501  
Db 1443 GAGCAG 1502  
QY 1502 GAAAGACTCGAAGCTTCTCTCAAGAGGCTTCCAAAGGCTGGGCTGCTTGTATATGA 1561  
Db 1503 GAAAGACTCGAAGCTTCTCTCAAGAGGCTTCCAAAGGCTGGGCTGCTTGTATATGA 1562  
QY 1562 CCGGTGTGTGGGGGAGAGAGAACTGGAAGTCTTATCAGTTCGGGGGCGTCTCGCA 1621  
Db 1563 CCGGTGTGTGGGGGAGAGAGAACTGGAAGTCTTATCAGTTCGGGGGCGTCTCGCA 1622  
QY 1622 TCAAGTGTGTGTGTGACAGTGAAGGCTTCCGAGATCAAAACAGTCCCAAGATTCGA 1681  
Db 1623 TCAAGTGTGTGTGTGACAGTGAAGGCTTCCGAGATCAAAACAGTCCCAAGATTCGA 1682  
QY 1682 TACAGATGATTAACCTGCCAAGAACAGATTTTAAAGGCGCAAAAATCTTTTCTTG 1741  
Db 1683 TACAGATGATTAACCTGCCAAGAACAGATTTTAAAGGCGCAAAAATCTTTTCTTG 1742

QY 1742 GGAGTCTACAAATTTGGAAATGAAAAAACCAGACATGATGTTTTTATTTATTTAT 1801  
Db 1743 GGAGTCTACAAATTTGGAAATGAAAAAACCAGACATGATGTTTTTATTTATTTAT 1802  
QY 1802 TATTTATGAAAGGTGTGATACCATTTATGATTTATGTGAAGGAGATGACACACCCAGCTTT 1861  
Db 1803 TATTTATGAAAGGTGTGATACCATTTATGATTTATGTGAAGGAGATGACACACCCAGCTTT 1862  
QY 1862 TGAAGGTGTGGGGGTGAAGATCTGAGCAGACCCCACTGGGAAACCAATCTGACCTGGCCC 1921  
Db 1863 TGAAGGTGTGGGGGTGAAGATCTGAGCAGACCCCACTGGGAAACCAATCTGACCTGGCCC 1922  
QY 1922 ATGGCTGTTTTCCCAAGATTCAGTTCTTGGAGGAAAGGCTCTGGCCCTGACTCCGCTGT 1981  
Db 1923 ATGGCTGTTTTCCCAAGATTCAGTTCTTGGAGGAAAGGCTCTGGCCCTGACTCCGCTGT 1982  
QY 1982 GTTCCGAGACACGTGTGACCGGACCGCCGCTGTGATTTCTTGGCTGGGTCTGGAG 2041  
Db 1983 GTTCCGAGACACGTGTGACCGGACCGCCGCTGTGATTTCTTGGCTGGGTCTGGAG 2042  
QY 2042 GTGTCTGTGAGACACCTGCTCCTCAACAGAGAGGTGAGCCATTTTGTGACATCCAGCT 2101  
Db 2043 GTGTCTGTGAGACACCTGCTCCTCAACAGAGAGGTGAGCCATTTTGTGACATCCAGCT 2102  
QY 2102 GAAACATGGGAAACAACCTGAAAAAGCAGGAGCTCCCGGATCAGGGAAGCTCTGCTGTGC 2161  
Db 2103 GAAACATGGGAAACAACCTGAAAAAGCAGGAGCTCCCGGATCAGGGAAGCTCTGCTGTGC 2162  
QY 2162 TGGCTTCCCATGACCACTCTCTCTGTAATTTATCTTGAATCTGAGACAGATTG 2221  
Db 2163 TGGCTTCCCATGACCACTCTCTCTGTAATTTATCTTGAATCTGAGACAGATTG 2222  
QY 2222 CCGGTTTATTAACCTGTTTTTATCTGAGAACAAACCGGTTTGGAAATTAATGCTGCTTTT 2281  
Db 2223 CCGGTTTATTAACCTGTTTTTATCTGAGAACAAACCGGTTTGGAAATTAATGCTGCTTTT 2282  
QY 2282 TTCCCACTCCCAAGAGCTCTCAAGCTCATTTCCAGCGGCCCCCTCGGCTTGGAGAGGTA 2341  
Db 2283 TTCCCACTCCCAAGAGCTCTCAAAATCATTTCCAGCGGCCCCCTCGGCTTGGAGAGGTA 2342  
QY 2342 GTGTAACTCCGATCCAGAGGCTTACCGCTGACAGAGTGGCTTCCCGTATCCCGGTGG 2401  
Db 2343 GTGTAACTCCGATCCAGAGGCTTACCGCTGACAGAGTGGCTTCCCGTATCCCGGTGG 2402  
QY 2402 AAAAGCCCTGACCAAGGAGGCTTGAAGTGGCTGTGTCCCTCAACCGCTGACACACC 2461  
Db 2403 AAAAGCCCTGACCAAGGAGGCTTGAAGTGGCTGTGTCCCTCAACCGCTGACACACC 2462  
QY 2462 CACCTTCAGAGTGCAGTGTGGGCAAGGAGCACTCAAGAGACAGGACAGGCGCTTGGC 2521  
Db 2463 CACCTTCAGAGTGCAGTGTGGGCAAGGAGCACTCAAGAGACAGGACAGGCGCTTGGC 2522  
QY 2522 AAGACATCAGACACACCAACCAAGGCTGGAACCCAGGCGCGGCTGTATCCAG 2581  
Db 2523 AAGACATCAGACACACCAACCAAGGCTGGAACCCAGGCGCGGCTGTATCCAG 2582  
QY 2582 CAGGTGACCTGCAAGTCCCGCTCTCTGAGGTCCAGGCTCTTCAACAGAAACCAAGGCG 2641  
Db 2583 CAGGTGACCTGCAAGTCCCGCTCTCTGAGGTCCAGGCTCTTCAACAGAAACCAAGGCG 2642  
QY 2642 CTGTGCTCCGAGGCTTCTCTTCAAGACCTTCTCTCAACGAGGCCACTTGGGATGCAAGATG 2701  
Db 2643 CTGTGCTCCGAGGCTTCTCTTCAAGACCTTCTCTCAACGAGGCCACTTGGGATGCAAGATG 2702  
QY 2702 CAGCGAGCTAGAGACCCCTCCACGAGCTGACCTCGGCTGCAATTAAGTTACGTGAGGC 2761  
Db 2703 CAGCGAGCTAGAGACCCCTCCACGAGCTGACCTCGGCTGCAATTAAGTTACGTGAGGC 2762  
QY 2762 CTGTCTCTCGGGGCTGGAAGTGGCAAGTTCATGATTGCTCTTGTGACCCCTTGGAGCAA 2821  
Db 2763 CTGTCTCTCGGGGCTGGAAGTGGCAAGTTCATGATTGCTCTTGTGACCCCTTGGAGCAA 2822

Oy 2822 GCGCGCAGAGTGTGCTGAGACAGCTGGCGGGGGGCCCCAAGCTGCGCGGCTC 2881  
 Db 2823 GCGCGCAGAGTGTGCTGAGACAGCTGGCGGGGGGCCCCAAGCTGCGCGGCTC 2882  
 Oy 2882 CAGCCCAACCCAGCTGTGTGCTGAGAGCTGAGCCCAACCTCCCGAGCACTGTATCTGAGTA 2941  
 Db 2883 CAGCCCAACCCAGCTGTGTGCTGAGAGCTGAGCCCAACCTCCCGAGCACTGTATCTGAGTA 2942  
 Oy 2942 ACGGCTAAGAACCTCTCTCTCTGTTTGAAGAACAGTGTGGGTGTCTCAATTCGTAA 3001  
 Db 2943 ACGGCTAAGAACCTCTCTCTCTGTTTGAAGAACAGTGTGGGTGTCTCAATTCGTAA 3002  
 Oy 3002 CATTCATCTCAATTTTAAAAAGGTTCCTGACGCGCCCGAGCGCCCGCGCGGTG 3061  
 Db 3003 CATTCATCTCAATTTT-AAAAAGGTTCCTGACGCGCCCGAGCGCCCGCGCGGTG 3061  
 Oy 3062 AGCGTGTGTGCTGAGAGCTGGGCGGCGGCTTCCCGTGGCGCTCTGCGCAGGTGCTT 3121  
 Db 3062 AGCGTGTGTGCTGAGAGCTGGGCGGCGGCTTCCCGTGGCGCTCTGCGCAGGTGCTT 3121  
 Oy 3122 CTGGGCAACCCATCTCTGCTGTTTCAATTTGACGTGACGTGACAGAGCACTCACACAA 3181  
 Db 3122 CTGGGCAACCCATCTCTGCTGTTTCAATTTGACGTGACGTGACAGAGCACTCACACAA 3181  
 Oy 3182 TAAACCTTCTGTAAGC 3199  
 Db 3182 TAAACCTTCTGTAAGC 3199

RESULT 4  
 AK001599  
 LOCUS  
 DEFINITION Homo sapiens cDNA FLJ10737 fis, clone NT2RP3001426, weakly similar to Dnaj PROTEIN.  
 ACCESSION AK001599  
 VERSION AK001599.1 GI:7022951  
 KEYWORDS oligo capping; fis (full insert sequence).  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
 AUTHORS  
 1 Isegai, T., Oka, T., Hayashi, K., Sugiyama, T., Otsuki, T., Suzuki, Y., Nishikawa, T., Nagai, K., Sugano, S., Shiratori, A., Sudo, H., Wagatsuma, M., Hosokita, T., Kaku, Y., Kodaira, H., Komdo, H., Sugawara, M., Takehashi, M., Chiba, Y., Ishida, S., Murakawa, K., Ono, Y., Takiguchi, S., Watanabe, S., Kimura, K., Murakami, K., Ishii, S., Kawai, Y., Saito, K., Yamamoto, J., Wakamatsu, A., Nakamura, Y., Nagahara, K., Masuko, Y., Ninomiya, K. and Iwayanagi, T.  
 NEDO human cDNA sequencing project  
 Unpublished  
 2 (bases 1 to 3199)  
 Isegai, T. and Otsuki, T.  
 Direct Submission  
 Submitted (16-FEB-2000) Takao Isegai, Helix Research Institute, Genomics Laboratory, 1532-3 Yana, Kisarazu, Chiba 229-0812, Japan (E-mail: genome@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)  
 NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing; Research Association for Biotechnology; cDNA library construction; 5'- & 3'-end one pass sequencing and clone selection; Helix Research Institute (supported by Japan Key Technology Center etc.) and Department of Virology, Institute of Medical Science, University of Tokyo.

# FEATURES source

1..3199  
 /location/Qualifiers  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="NT2RP3001426"  
 /cell\_line="NT2"  
 /cell\_type="teratocarcinoma"  
 /clone\_lib="NT2RP3"

/note="cloning vector: pMR18SFL3-mRNA from NT2 neuronal precursor cells after 2-weeks retinotic acid (RA) induction."  
 16..1695  
 /note="unnamed protein product"  
 /codon\_start=1  
 /protein\_id="BA91780.1"  
 /db\_xref="GI:7022952"  
 /translation="MATPAUSEEELDNEEDYSLNVRREASSEELKAYRILCMYHPD KHRDPEIKSOERLFDNLVHQAYEVLSDPQTRALYDYGKGLMEGEVERRTPEK IREPERLQREERERLQERTNPKGTISVGDATDLDPRDEYEDVSGSPQIEIN KHIISOSIEAPLATDPAIISGLSLONGGCGSINPALRVTASAGMGSLFEGAGL OGLFGLKLEPRLNLPFCFVTTNCAIOPSSRGIRPGKFTVLARNLKNTMGVYLQMRG I OSAMTSIYADPTSHPTVALOGIHSPLIYOKRPODDOTRYVGSIKAPRGV VETGAERKISRHSVLGAANSVGPQGVSLKYKLNRSQTFPPIHLTDQILPSMPTA TVSPVLVYFAMRLLIKPYLRACKERLEKORSAAVTDVLOKQOASAVRLQOESVR RIIEAESRNGLIIVANWGFVNDKRSKEXKVIDVYPLQCLVDSKLITLAEASK AGHPGFYDPCVGEERKILKVLXQPRGVHLQVMDSEALRIPKQSHRIDTGS"

## ORIGIN

Query Match 98.7%; Score 3171.6; DB 9; Length 3199;  
 Best Local Similarity 99.7%; Pred. No. 0;  
 Matches 3188; Conservative 0; Mismatches 9; Indels 1; Gaps 1;  
 Oy 2 AAAGTTGGCAAGATGGCGAAGCGCTTGAGCGAGGAGAGCTGGACAAATGAAGCTATTTA 61  
 Db 3 AAAGTTGGCAAGATGGCGAAGCGCTTGAGCGAGGAGAGCTGGACAAATGAAGCTATTTA 62  
 Oy 62 CTGCTGCTGACAGCTGGCGAGGAGGCGCTTCTTGAAGAGCTGAAAGCTCTTACCGGAG 121  
 Db 63 CTGCTGCTGACAGCTGGCGAGGAGGCGCTTCTTGAAGAGCTGAAAGCTCTTACCGGAG 122  
 Oy 122 GCTGTGTATGCTTACCATTCACAGCAACAGACAGACAGCTCAAGTCAAGCGCGA 181  
 Db 123 GCTGTGTATGCTTACCATTCACAGCAACAGACAGCTCAAGTCAAGCGCGA 182  
 Oy 182 ACAGCTGTTTAACTTGTGTACAGAGCTTGAAGGCTTAGTGAACCCCAACCAAGGCG 241  
 Db 183 ACAGCTGTTTAACTTGTGTACAGAGCTTGAAGGCTTAGTGAACCCCAACCAAGGCG 242  
 Oy 242 CATCTATGATATATATGGAAGAGAGAGCTGAAATGGAAGATGGAGGTTGTGAAG 301  
 Db 243 CATCTATGATATATATGGAAGAGAGAGCTGAAATGGAAGATGGAGGTTGTGAAG 302  
 Oy 302 GAGGAGAACCCCTGCTGAAATTTGAGAGAGATTGAGCGGCTGACAGAGAGAGAGA 361  
 Db 303 GAGGAGAACCCCTGCTGAAATTTGAGAGAGATTGAGCGGCTGACAGAGAGAGAGA 362  
 Oy 362 GAGGAGATTCAGAGAGAACCAATCCCAAGGGAAGATGAGGTTGAGATGACAC 421  
 Db 363 GAGGAGATTCAGAGAGAACCAATCCCAAGGGAAGATGAGGTTGAGATGACAC 422  
 Oy 422 CGACCTTTTGTATGCTATGATGAGAGATGGAAGATGTCTCCGAGTAGCTTTCCGCA 481  
 Db 423 CGACCTTTTGTATGCTATGATGAGAGATGGAAGATGTCTCCGAGTAGCTTTCCGCA 482  
 Oy 482 GATTGAAATTAATTAATGACATATCCAGTCCATTTAGAGCACTTTGAACAGGACAGA 541  
 Db 483 GATTGAAATTAATTAATGACATATCCAGTCCATTTAGAGCACTTTGAACAGGACAGA 542  
 Oy 542 CACAGCCATCTCTTGGAGAGCTTCTCAACCCAGAAATGGAATGAGAGAGGTTCCATTAA 601  
 Db 543 CACAGCCATCTCTTGGAGAGCTTCTCAACCCAGAAATGGAATGAGAGAGGTTCCATTAA 602  
 Oy 602 CTTTGCGCTCAGACAGATTACTTCGCAAGAGAGATGGAGAGATTGGAATTTGAGCTGG 661  
 Db 603 CTTTGCGCTCAGACAGATTACTTCGCAAGAGAGATGGAGAGATTGGAATTTGAGCTGG 662  
 Oy 662 AGACTTACAGAGGCGCTTTGTGGGTCTCAAGCTGTTCCGTAATCTCAACCAAGATGCTT 721  
 Db 663 AGACTTACAGAGGCGCTTTGTGGGTCTCAAGCTGTTCCGTAATCTCAACCAAGATGCTT 722  
 Oy 722 TGTGACAAACAACCTGTCTCTGACAGTTTATCCCGTGAATCCGACCCGCGCTGACAC 781

|||||  
Db 723 TGTGACAACTGCTCTGACAGTTTTCATCCCGTGAATCCGACCCGCGCTGACAC 782  
|||||  
QY 782 TGTCTAGAGCTGGAACTAGACAAAGAACACCGTGGGTACTCTGACAGTGGGATGGGGTAT 841  
783 TGTCTAGAGCTGGAACTAGACAAAGAACACCGTGGGTACTCTGACAGTGGGATGGGGTAT 842  
QY 842 CCAGTCAGCCATGAACACTAGATCGTCCGAGACACTAAACAGCCACTTCACTGTGGC 901  
843 CCAGTCAGCCATGAACACTAGATCGTCCGAGACACTAAACAGCCACTTCACTGTGGC 902  
Db 902 CTTGCAAGCTGGGAATCCCTCACTCTCTTGGCATGTATGACTATGACACAAATTTCCAA 961  
903 CTTGCAAGCTGGGAATCCCTCACTCTCTTGGCATGTATGACTATGACACAAATTTCCAA 962  
QY 962 TGAAGATCAGACTCGTGTGAAGAGATCCCTCAAGAGAGGCTTCTTGGGAGCGGTGGTGA 1021  
963 TGAAGATCAGACTCGTGTGAAGAGATCCCTCAAGAGAGGCTTCTTGGGAGCGGTGGTGA 1022  
Db 1022 GTACGAGCTGAGAGAGAGATCTCCAGGCAACGCGTTTGGGTGACAGCTGTGACGCTTGG 1081  
1023 GTACGAGCTGAGAGAGAGATCTCCAGGCAACGCGTTTGGGTGACAGCTGTGACGCGTTGG 1082  
QY 1082 AGTTTCCACAGGGGCTTTCTCTCAAAGTCAAGCTCAACAGGCGCATGACATATCTTCTT 1141  
1083 AGTTTCCACAGGGGCTTTCTCTCAAAGTCAAGCTCAACAGGCGCATGACATATCTTCTT 1142  
Db 1142 CCCATTTCACTTGAACGAGACGAGCTTCCGACAGGCGCATGTCTATGCGACCGGTGGGCGC 1201  
1143 CCCATTTCACTTGAACGAGACGAGCTTCCGACAGGCGCATGTCTATGCGACCGGTGGGCGC 1202  
QY 1202 TCTAGTGTCTACTTTGCTCATGACCGGTCTGATCATCAACCATACCTCAGGGCTCAGAA 1261  
1203 TCTAGTGTCTACTTTGCTCATGACCGGTCTGATCATCAACCATACCTCAGGGCTCAGAA 1262  
Db 1262 AGAAGAGGAATTGGAG 1321  
1263 AGAAGAGGAATTGGAG 1322  
QY 1322 AGAGGCGGAGTCCGCTGTCCGCGTGAATGACGAATCTGTCCGAAGATTAATTGAGGAG 1381  
1323 AGAGGCGGAGTCCGCTGTCCGCGTGAATGACGAATCTGTCCGAAGATTAATTGAGGAG 1382  
Db 1382 AGAGTCCAGATGGGCTCTCATATCGTCAATGCTGCTGTAACGGAAGTTTGTCAATGACA 1441  
1383 AGAGTCCAGATGGGCTCTCATATCGTCAATGCTGCTGTAACGGAAGTTTGTCAATGACA 1442  
QY 1442 GAGCAG 1501  
1443 GAGCAG 1502  
Db 1502 GAGAGACTCGAGCTCATCTCCACGAGGCTCCAAAGGCTGGGCTGCTGGCTTTATGA 1561  
1503 GAGAGACTCGAGCTCATCTCCACGAGGCTCCAAAGGCTGGGCTGCTGGCTTTATGA 1562  
QY 1562 CCCGTGTGTGGGAG 1621  
1563 CCCGTGTGTGGGAG 1622  
Db 1622 TCAAGT 1681  
1623 TCAAGT 1682  
QY 1682 TACAGATGATTAAGCTGCAAGAGATTTTAAAGGCGCAAAATCTTTTCTG 1741  
1683 TACAGATGATTAAGCTGCAAGAGATTTTAAAGGCGCAAAATCTTTTCTG 1742  
Db 1742 GAGGTCTACAAATTTGGAAATGAAAAAACCAGACATCAGATGTTTATTTATATAT 1801  
1743 GAGGTCTACAAATTTGGAAATGAAAAAACCAGACATCAGATGTTTATTTATTTAT 1802  
QY 1802 TATTTATGAAGGT 1861  
|||||

Db 1803 TATTTATGAAGGT 1862  
QY 1862 TGAAGGTGTGTGGGGTAGAGACTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1921  
1863 TGAAGGTGTGTGGGGTAGAGACTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1922  
QY 1922 ATGGCTGTTTTCCAAAGATCAGTTCTGTGAAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 1981  
1923 ATGGCTGTTTTCCAAAGATCAGTTCTGTGAAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 1982  
Db 1982 GTTCCGAGACACGTGTCTGACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2041  
1983 GTTCCGAGACACGTGTCTGACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2042  
QY 2042 GTGTCTGTGAAGACCTTGCCTCTCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2101  
2043 GTGTCTGTGAAGACCTTGCCTCTCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2102  
Db 2102 GAAATGGGAAACAACTGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2161  
2103 GAAATGGGAAACAACTGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2162  
QY 2162 TGGCTTCCATGACACCTCTCTCTGTGTGAATATTAATCTGTGTGAATCTGTGAGAGAT 2221  
2163 TGGCTTCCATGACACCTCTCTCTGTGTGAATATTAATCTGTGTGAATCTGTGAGAGAT 2222  
Db 2222 CGGGTTTATTAATCTGTTTTATCTGAGAAACAAACGGGTTTGAATTAATGATGCTTTT 2281  
2223 CGGGTTTATTAATCTGTTTTATCTGAGAAACAAACGGGTTTGAATTAATGATGCTTTT 2282  
QY 2282 TTCCCACTTCCAGAGCTGTCAAGTCAATTCACAGGAGAGAGAGAGAGAGAGAGAGAG 2341  
2283 TTCCCACTTCCAGAGCTGTCAAAATCAATTCACAGGAGAGAGAGAGAGAGAGAGAGAG 2342  
Db 2342 GTGTAACTCCGATCCAG 2401  
2343 GTGTAACTCCGATCCAG 2402  
QY 2402 AAAAGCGCTGCAACAG 2461  
2403 AAAAGCGCTGCAACAG 2462  
Db 2462 CACTTCAGAGTGCAGTGTGTGGCAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2521  
2463 CACTTCAGAGTGCAGTGTGTGGCAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2522  
QY 2522 AAGACATCAGACACACCCAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2581  
2523 AAGACATCAGACACACCCAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2582  
Db 2582 CAGGTGACACTGAGTCCCGCTCTCTGAGAGTCAAGAGTCTCTCAAGAGACACAGAGAG 2641  
2583 CAGGTGACACTGAGTCCCGCTCTCTGAGAGTCAAGAGTCTCTCAAGAGACACAGAGAG 2642  
QY 2642 CTGTGTCTGGAGAGCTTCTCTCAAGAGTCTCTCAAGAGTCTCTCAAGAGTCTCTCAAG 2701  
2643 CTGTGTCTGGAGAGCTTCTCTCAAGAGTCTCTCAAGAGTCTCTCAAGAGTCTCTCAAG 2702  
Db 2702 CAGCGAGGTAGAGACCCCTCTCAAGAGTCTCTCAAGAGTCTCTCAAGAGTCTCTCAAG 2761  
2703 CAGCGAGGTAGAGACCCCTCTCAAGAGTCTCTCAAGAGTCTCTCAAGAGTCTCTCAAG 2762  
QY 2762 CTGTCTCTGGGGCTGTGAAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2821  
2763 CTGTCTCTGGGGCTGTGAAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2822  
Db 2822 GCGCGGACAGGT 2881  
2823 GCGCGGACAGGT 2882  
QY 2882 CAGCCACCCACAGCTGT 2941  
2883 CAGCCACCCACAGCTGT 2942  
|||||



QY	2942	ACGGTAGAAACCTCCCTCCTCTGGATTGTGAAAAAGCACTGGGGTGTGCAATTCTGTAA	3001
Db	2943	ACGGTAGAAACCTCCCTCCTCTGGATTGTGAAAAAGCACTGGGGTGTGCAATTCTGTAA	3002
QY	3002	CATTCACTCCATTTTTTAAAAAGGTTCTCTGACGGACCCACAGGACCGAGCGGGTGG	3061
Db	3003	CATTCACTCCATTTTTT - AAAAAGGTTCTCTGACGGACCCACAGGACCGAGCGGGTGG	3061
QY	3062	AGCGTCGTGTGTCATGAGCTCGGGCCCCGGGCTTCCGTCGGGCTCTGCGGAGGTGCTT	3121
Db	3062	AGCGTCGTGTGTCATGAGCTCGGGCCCCGGGCTTCCGTCGGGCTCTGCGGAGGTGCTT	3121
QY	3122	CTGGGACCCCACTCCCTCGCGTTCACTTTGACAGTGCAGTGCAGTAACAGAGGCACTGCACCAA	3181
Db	3122	CTGGGACCCCACTCCCTCGCGTTCACTTTGACAGTGCAGTGCAGTAACAGAGGCACTGCACCAA	3181
QY	3182	TAAACCTTTCTGAAAGC	3199
Db	3182	TAAACCTTTCTGAAAGC	3199

RESULT 5					
HSM805121					
LOCUS	HSM805121	3178 bp	mRNA	linear	PRI 12-JUL-2002
DEFINITION	Homo sapiens mRNA; CDNA	DKFZp727Ct81	(from	clone	DKFZp727Ct81).
ACCESSION	AL833841				
VERSION	AL833841.1	GI:21739308			
KEYWORDS					
SOURCE	Homo sapiens				
ORGANISM	Homo sapiens (human)				

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

1. (baes 1 to 3178)  
Poustka, A., Klein, M., Mewes, H.W., Gassenhuber, J., and Wiemann, S.  
Direct Submission  
Submitted (10-JUN-2002) GSF, Institut fuer Bioinformatik MPS,  
Ingolstaedter Landstr. 1, 85764 Neuberg, Germany  
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer

**FEATURES**  
 Heidelberg/Germany) within the cDNA sequencing consortium of the German Genome Project. This clone (DKFZp27C181) is available at the RZPD in Berlin. Please contact the RZPD: Resourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further information about the clone and the sequencing project is available at <http://mips.gsf.de/proj/cDNA/>.  
 Location/Qualifiers

```

1. .3178
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/map="p36.21-36.33"
/clone="DKFzp727C181"
/tissue_type="breast cancer"
/clone_idb="727 (synonym: hmcfl) . Vector pSport1; host
DH10B; sites NotI + SalI"
/dev_stage="adult"
1. .3178
gene
/gene="DKFzp727C181"
CDS
<1. .1660

```

```

/gene="DKFZ2727C181"
/notes="weak similarity to putative DNAJ protein (A.
chaliana)"
/codon_start=2
/product="hypothetical protein"
/protein_id="CAD36701.1"
/db_xref="GI:21739309"
/db_xref="SPTREMBL:Q8NDM2"
/translation="EELDNEEDYSLILNVRREASSEELKAAYRRLCMLYHPDKHDPDL
KQAEKRLFLVHQAEVGLSDPQRAIYIYKRLGLEMGWGVRRRTPTIKRIIEEFER
LQREBEERLFOAVTNPQKGISVGDATDLFPRYEEYEDVSGSFPQIEIKMHI SOS

```

IEMPLATPTALLSGSLSTONGNGGSGINPLAARVTSACQWELFEFGADJOGPLPG  
 KLRNLTUPRCFVTTTNCALQFSSRGIRPGLTTLVARNIDKNTVTGLDRNGTIOGAMTGS  
 IVDTDKTSHTFVALQDGIPIHSFALISTQHKFQDDDRKVGSKLACGPFCTVSYGMR  
 KIRSHSVLIGAAVSVGVQFQSGELVKYKLNRAQYTFPPIHLTDQLPSAMPATVPLV  
 YFPMHRLIIRPYDRAQKEKELQORBSAADVDYQKQZESAVRLMOESVRIITEEB  
 SRMGILIVNAYGKFPVDRKSEKRYVIDVTPVQLCLVDSKLIITFASKALPGFY  
 DPCGGEKRNKLVYQFPGVLHQMVLDSALKLRIPKQSHRIDTGG"  
 3145. 3150  
 poly\_a\_signal  
 /gene="DKFzp727Ct81"  
 3159  
 poly\_a\_site  
 /gene="DKFzp727Ct81"  
 ORIGIN

Query Match	98.3%	Score 3157.4	DB 9	Length 3178	
Best Local Similarity	99.8%	Pred. No. 0			
Matches 3172	Conservative	0	Mismatches 6	Indels 1	Gaps 1
QY	35	GGAGAGCTGGACATGAGACTATTACTGTGCTGGAACGTGCGCAGGAGGCTCTTC	94		
DB	1	GGAGAGCTGGACATGAGACTATTACTGTGCTGGAACGTGCGCAGGAGGCTCTTC	60		
QY	95	TGAAGACTGAAAGCTGCTTACCGGAGGCTCTGTATGCTTACCATTCAGACAAACAG	154		
DB	61	TGAAGACTGAAAGCTGCTTACCGGAGGCTCTGTATGCTTACCATTCAGACAAACAG	120		
QY	155	AGACCCGAGGTCAGTCAGACGAGGAGAAAGCATGTTTAACTTGTTCAACAGGCTTATGA	214		
DB	121	AGACCCGAGGTCAGTCAGTCAGGAGGAAACATGTTTAACTTGTTCAACAGGCTTATGA	180		
QY	215	AGTCTTATGATGACCCCAAAACAAGGCTATCTATGATATATATGGAAAGAGACTGGA	274		
DB	181	AGTCTTATGATGACCCCAAAACAAGGCTATCTATGATATATATGGAAAGAGACTGGA	240		
QY	275	AATGGAAGATATGGAGGCTTGTGAAAAGSAGAAACCCCTGCTGAAATTCGAGAGATT	334		
DB	241	AATGGAAGATATGGAGGCTTGTGAAAAGSAGAAACCCCTGCTGAAATTCGAGAGATT	300		
QY	335	TGAGCGGCTGCAGAGAGAGAGAAAGAGAGATTGCAGCAGCAACCAATCCCAAGG	394		
DB	301	TGAGCGGCTGCAGAGAGAGAGAAAGAGAGATTGCAGCAGCAACCAATCCCAAGG	360		
QY	395	AACGATCAGCGTTGGAGTAGATGCCACGACTTTTGTATGCTATGATGAGAGATGA	454		
DB	361	AACGATCAGCGTTGGAGTAGATGCCACGACTTTTGTATGCTATGATGAGAGATGA	420		
QY	455	AGATGTCCTCGGACATGACTTTCGACAGATTGAATTAAATAAATGACATATCCAGTC	514		
DB	421	AGATGTCCTCGGACATGACTTTCGACAGATTGAATTAAATAAATGACATATCCAGTC	480		
QY	515	CATTGAGGCAACCTTGCACAGCAGACAGACAGGCAATCTCTCGAAGGCTCTCAACCA	574		
DB	481	CATTGAGGCAACCTTGCACAGCAGACAGACAGGCAATCTCTCGAAGGCTCTCAACCA	540		
QY	575	GAATGAAATAGAGAGAGGTTCTATTAACTTTGGCTCAGACGATTACTTCCGCAAGG	634		
DB	541	GAATGAAATAGAGAGAGGTTCTATTAACTTTGGCTCAGACGATTACTTCCGCAAGG	600		
QY	635	ATGGGAGAGATTGGAATTTGAGCTGAGACTTAAGAGGGCTTTGTTCCGCTCAAGCT	694		
DB	601	ATGGGAGAGATTGGAATTTGAGCTGAGACTTAAGAGGGCTTTGTTCCGCTCAAGCT	660		
QY	695	GTTCCGTAATCTCACACCAAGATGCTTTGTGAACAACACTGTGCTCTGCAATTTTCATC	754		
DB	661	GTTCCGTAATCTCACACCAAGATGCTTTGTGAACAACACTGTGCTCTGCAATTTTCATC	720		
QY	755	CCGTAAGATTCGAAACCGGCTGACACTGTGCTTGAACAACACTGTGCTCTGCAATTTTCATC	814		
DB	721	CCGTAAGATTCGAAACCGGCTGACACTGTGCTTGAACAACACTGTGCTCTGCAATTTTCATC	780		
QY	815	GAGCTACCTCAGATGGAGATGGAGATTCACAGTACGACCATGAACCTATGCTGCGAGA	874		
DB	781	GAGCTACCTCAGATGGAGATGGAGATTCACAGTACGACCATGAACCTATGCTGCGAGA	840		

QY 875 CACTAAAAACGACCTTCACTGTGGCCCTGACGCTGGGAATCCCTCACTCTTTTGACT 934  
Db 841 CACTAAAAACGACCACTTCACTGTGGCCCTGACGCTGGGAATCCCTCACTCTTTTGACT 900  
QY 935 GATCAGCTATCAGACAAATTTCCAAAGTGAAGATCAGATCTCGTGAAGATCCCTCA 994  
Db 901 GATCAGCTATCAGACAAATTTCCAAAGTGAAGATCAGATCTCGTGAAGATCCCTCA 960  
QY 995 AGCAGCTTCTTTGGACGCTGTGAGTACGAGCTGAGAGGAAGATCTCCAGGACAG 1054  
Db 961 AGCAGCTTCTTTGGACGCTGTGAGTACGAGCTGAGAGGAAGATCTCCAGGACAG 1020  
QY 1055 CGTTTGGGTGACAGCTGTACGCTTGAGTTCCACAGGCGCTTCTCTCAAGTCAAGCT 1114  
Db 1021 CGTTTGGGTGACAGCTGTACGCTTGAGTTCCACAGGCGCTTCTCTCAAGTCAAGCT 1080  
QY 1115 CAACAGGCGCAGTCAACATCTTCTCCCTATTCATTTGAAGGACCAAGTTCTGCCCAG 1174  
Db 1081 CAACAGGCGCAGTCAACATCTTCTCCCTATTCATTTGAAGGACCAAGTTCTGCCCAG 1140  
QY 1175 CGCATGTTCTATGCCACCGTGGGGCTCTAGTGTCTACTTTGGCATGTGACCGTCTGAT 1234  
Db 1141 CGCATGTTCTATGCCACCGTGGGGCTCTAGTGTCTACTTTGGCATGTGACCGTCTGAT 1200  
QY 1235 CATCAAAACATACCTCAGGCTCAGAAAAGAAAGAAATTGAGAAAGCAGAGGAAACCGC 1294  
Db 1201 CATCAAAACATACCTCAGGCTCAGAAAAGAAAGAAATTGAGAAAGCAGAGGAAACCGC 1260  
QY 1295 CGCCACCGATGTGTGAGAAAGAAAGAGGCGGAGATCCGCTGTCCGGGTGATGACAGA 1354  
Db 1261 CGCCACCGATGTGTGAGAAAGAAAGAGGCGGAGATCCGCTGTCCGGGTGATGACAGA 1320  
QY 1355 ATCTGTCCGAAGATTAATTGAGGACGAGAGATCCAGAAATGGGCTCATCTGATCAATGC 1414  
Db 1321 ATCTGTCCGAAGATTAATTGAGGACGAGAGATCCAGAAATGGGCTCATCTGATCAATGC 1380  
QY 1415 CTGGTACGGAAGTTGTCAATGAACAAGACAGAGAGCGAAGAGTGAAGTGAATGA 1474  
Db 1381 CTGGTACGGAAGTTGTCAATGAACAAGACAGAGAGCGAAGAGTGAAGTGAATGA 1440  
QY 1475 CGTGAAGTGGCCCTGAGAGGCGCTGAGAGACTCGAAGTCACTCTCAAGAGCGCTC 1534  
Db 1441 CGTGAAGTGGCCCTGAGAGGCGCTGAGAGACTCGAAGTCACTCTCAAGAGCGCTC 1500  
QY 1535 CAAGGCTGGGCTGCTGGCTTTTATGACCCGTGTGTGGGGAAGAGAAAGATCTGAAAT 1594  
Db 1501 CAAGGCTGGGCTGCTGGCTTTTATGACCCGTGTGTGGGGAAGAGAAAGATCTGAAAT 1560  
QY 1595 GCTCTATCAGTTCCGGGGCGTCTGCAATCAGTGAATGTGTCTGACAGTGAAGCCCTCCG 1654  
Db 1561 GCTCTATCAGTTCCGGGGCGTCTGCAATCAGTGAATGTGTGTCTGACAGTGAAGCCCTCCG 1620  
QY 1655 GATACCAAGCAAGTCCCAAGATCCGATCAGATGATAAATGCCAAGAACCAAGTTT 1714  
Db 1621 GATACCAAGCAAGTCCCAAGATCCGATCAGATGATAAATGCCAAGAACCAAGTTT 1680  
QY 1715 TAAAAAGCCCAAAAAATCTTTTCTGGAGTCTACAATTTTGGAAATGAAAAACCCAG 1774  
Db 1681 TAAAAAGCCCAAAAAATCTTTTCTGGAGTCTACAATTTTGGAAATGAAAAACCCAG 1740  
QY 1775 ACATCAGATGTTTTTATTTATTTATTTATTTATTAAGAGTGTGACATTTCAATTAATGT 1834  
Db 1741 ACATCAGATGTTTTTATTTATTTATTTATTTATTAAGAGTGTGACATTTCAATTAATGT 1800  
QY 1835 GAAGGGACATGACGACCCCAAGTTTGAAGGCTGTGGGGGTAGAGCTGAGGACAGCC 1894  
Db 1801 GAAGGGACATGACGACCCCAAGTTTGAAGGCTGTGGGGGTAGAGCTGAGGACAGCC 1860  
QY 1895 ACTGGGAACGAGATGAGGCTGGCCCATGGCTGTTTCCCAAGATCAGTCTTGAGG 1954  
Db 1861 ACTGGGAACGAGATGAGGCTGGCCCATGGCTGTTTCCCAAGATCAGTCTTGAGG 1920  
QY 1955 GAAGGCTCTGGCCCTGACTCCGCTGTCTCCAGACACAGTGTGAACCGACCCGCG 2014

Db 1921 GAAGGCTCTGGCCCTGACTCCGCTGTGTCCGAGACACAGTGTGACCGACCCGCGCG 1980  
QY 2015 CCTGTAGTTCTTTGGTGGGTGAGAGTGTCTGTGAGACACACCTGCCCTCCACAGGA 2074  
Db 1981 CCTGTAGTTCTTTGGTGGGTGAGAGTGTCTGTGAGACACACCTGCCCTCCACAGGA 2040  
QY 2075 GCGTGAACCACTTCTGCACTGCAAGCTGAACATGGGAAACAACTGAAAGAGGACGCG 2134  
Db 2041 GCGTGAACCACTTCTGCACTGCAAGCTGAACATGGGAAACAACTGAAAGAGGACGCG 2100  
QY 2135 CTCGCCGTGAGGAGCTTGTCTGTGTGCTTCCATGACCACTCTCTCTGTAAT 2194  
Db 2101 CTCGCCGTGAGGAGCTTGTCTGTGTGCTTCCATGACCACTCTCTCTGTAAT 2160  
QY 2195 ATTACTGCTGAATCTGAGACGATTTGGGGTTTAAACGCTTTTATCTGAGAAC 2254  
Db 2161 ATTACTGCTGAATCTGAGACGATTTGGGGTTTAAACGCTTTTATCTGAGAAC 2220  
QY 2255 AACGGTTTGAATAATGATCGTCTTTTCCCACTCCAGAGCTGCTCAAGTCAATCCA 2314  
Db 2221 AACGGTTTGAATAATGATCGTCTTTTCCCACTCCAGAGCTGCTCAATCAATCCA 2280  
QY 2315 CGGCCCCCTCGGCTTTGGGACAGGGTATGTAACTCCGATCCAGGGCTTAGCCTGAC 2374  
Db 2281 CGGCCCCCTCGGCTTTGGGACAGGGTATGTAACTCCGATCCAGGGCTTAGCCTGAC 2340  
QY 2375 ACAGGTGGCTTCCGATATCCGGTGGGAAAGCCCTGACACAGGGCTTAGGCTGAC 2434  
Db 2341 ACAGGTGGCTTCCGATATCCGGTGGGAAAGCCCTGACACAGGGCTTAGGCTGAC 2400  
QY 2435 CTGTGTCCCTCCACCGCTGACCAACCACTCCAGAGTGCAGTGTGGGCAAGGACAC 2494  
Db 2401 CTGTGTCCCTCCACCGCTGACCAACCACTCCAGAGTGCAGTGTGGGCAAGGACAC 2460  
QY 2495 TCAAGAGACAGACCAAGGCGCTTGGCAAGATCAGACACACCAACCAAGGCGTGG 2554  
Db 2461 TCAAGAGACAGACCAAGGCGCTTGGCAAGATCAGACACACCAACCAAGGCGTGG 2520  
QY 2555 ACCCCAGGCGCGGCGGTGTAACCAAGAGTGGACCTGACAGCTCCGCTCTGACAGT 2614  
Db 2521 ACCCCAGGCGCGGCGGTGTAACCAAGAGTGGACCTGACAGCTCCGCTCTGACAGT 2580  
QY 2615 CCAGCTCTCTCAAGAGAACCAAGGCGTGTCTCCGAGCCCTTCTTGACACCTTCTCT 2674  
Db 2581 CCAGCTCTCTCAAGAGAACCAAGGCGTGTCTCCGAGCCCTTCTTGACACCTTCTCT 2640  
QY 2675 CCACTGTCCCACTTTGGGATGCAAGATGACAGGGAAGTAAAGCCCTCCAGGCTTGAC 2734  
Db 2641 CCACTGTCCCACTTTGGGATGCAAGATGACAGGGAAGTAAAGCCCTCCAGGCTTGAC 2700  
QY 2735 CTCGCTGAGTAAAGTTACGTGAAGCCGTCTCTCGGGGCGTGAAGTGGAGGCATCA 2794  
Db 2701 CTCGCTGAGTAAAGTTACGTGAAGCCGTCTCTCGGGGCGTGAAGTGGAGGCATCA 2760  
QY 2795 GTTGTCTTGTGACCCCTCCGAGCAAGCGCCGCAAGGTGTGTGAGACAGCTGAGCG 2854  
Db 2761 GTTGTCTTGTGACCCCTCCGAGCAAGCGCCGCAAGGTGTGTGAGACAGCTGAGCG 2820  
QY 2855 CGGGGGGCCCAAGCTGCGCGGCTCCAGCCCAACCAAGCTGTGTGAGAGTCAAGCC 2914  
Db 2821 CGGGGGGCCCAAGCTGCGCGGCTCCAGCCCAACCAAGCTGTGTGAGAGTCAAGCC 2880  
QY 2915 AACCTCCCAAGACGTGTATCTGAGTAAGGCTGAAGAACCTGCTCTGTGTTTGA 2974  
Db 2881 AACCTCCCAAGACGTGTATCTGAGTAAGGCTGAAGAACCTGCTCTGTGTTTGA 2940  
QY 2975 AGCAGTTCCGGTGTGTCATTTCTGTAACTGTAACTCAATTTTAAAAAGTTTCTCT 3034  
Db 2941 AGCAGTTCCGGTGTGTCATTTCTGTAACTGTAACTCAATTTTAAAAAGTTTCTCT 2999  
QY 3035 GAAGGCCCAAGGCGCGGAGCGCGGTGAGCGTGTGTGACATGAGCTGTGGCCCGGCT 3094



Db 3000 GACGGCCCCAGCGCCGAGCGGTGAGCGTGTGTTGATGAGACTGCGGCCCGCGGCT 3059  
Qy 3095 TCCCGTGGCCTTTCGCCAGGTGTTCTGGGACACCATCTCTGCGTTTCATTGGCGT 3154  
Db 3060 TCCCGTGGCCTTTCGCCAGGTGTTCTGGGACACCATCTCTGCGTTTCATTGGCGT 3119  
Qy 3155 CGACTGTACAGAGGACCTCACCATTAACCTTCTGGAAGCAGAAAAA 3213  
Db 3120 CGACTGTACAGAGGACCTCACCATTAACCTTCTGGAAGCAGAAAAA 3178

RESULT 6  
AK025491 2064 bp mRNA linear PRI 13-SEP-2003  
LOCUS Homo sapiens cDNA: FLJ21838 f1s, clone HEP01752, highly similar to  
DEFINITION AK025491 Homo sapiens clone 24616 mRNA sequence.  
ACCESSION AK025491.1 GI:10438023  
VERSION AK025491.1  
KEYWORDS oligo capping; f1s (full insert sequence).  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
AUTHORS Kawabata, A., Hikiji, T., Kobatake, N., Inagaki, H., Ikema, Y.,  
Okamoto, S., Okitani, R., Ota, T., Suzuki, Y., Ohyashi, M., Nishi, T.,  
Shibahara, T., Tanaka, T., Nakamura, Y., Isogai, T. and Sugano, S.  
TITLE NEDO human cDNA sequencing project  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 2064)  
AUTHORS Sugano, S., Suzuki, Y., Ota, T., Ohyashi, M., Nishi, T.,  
Shibahara, T., Tanaka, T. and Nakamura, Y.  
TITLE Direct Submission  
JOURNAL Submitted (29-AUG-2000) Sumio Sugano, Institute of Medical Science,  
University of Tokyo, Laboratory of Genome Structure Analysis, Human  
Genome Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639,  
Japan (E-mail: f1cdna@ims.u-tokyo.ac.jp, Tel: 81-3-5449-5286,  
Fax: 81-3-5449-5416)  
COMMENT NEDO human cDNA sequencing project supported by Ministry of  
International Trade and Industry of Japan; cDNA full insert  
sequencing: Research Association for Biotechnology; cDNA library  
construction: 5' - & 3' - end one pass sequencing; Department of  
Virology and Human Genome Center, Institute of Medical Science,  
University of Tokyo (partly supported by Science and Technology  
Agency).

FEATURES  
source  
1. .2064  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="HEP01752"  
/cell\_line="HepG2"  
/cell\_type="hepatoma"  
/clone\_id="HEP"  
/note="Cloning vector pME186FL3"  
misc\_feature  
1. .2064  
/note="highly similar to AF052158 Homo sapiens clone 24616  
mRNA sequence"

ORIGIN  
Query Match 60.6%; Score 1947; DB 9; Length 2064;  
Best Local Similarity 99.6%; Pred. No. 0;  
Matches 1972; Conservative 0; Mismatches 5; Indels 2; Gaps 2;

Qy 1236 ATCAAAACCATACCTCAGGGCTCAGAAAGAAAGATTGAGAAAGCAGAGGAAAGCGCC 1295  
Db 2 AACAAACCATACCTCAGGGCTCAGAAAGAAAGATTGAGAAAGCAGAGGAAAGCGCC 61  
Qy 1296 GCCACCATGTCCTGACAGAAAGCAGAGGCGAGTCCGCTGTCGGCTATGACAGAA 1355  
Db 62 GCCACCATGTCCTGACAGAAAGCAGAGGCGAGTCCGCTGTCGGCTATGACAGAA 121  
Qy 1356 TCTGTCCGAAGATTAATTGAGGACAGAAAGTCCAGAAATGGGCTCATGTCAATGCC 1415

Db 122 TCTGTCCGAAGATTAATTGAGGACAGAAAGTCCAGAAATGGGCTCATGTCAATGCC 181  
Qy 1416 TGATACGAGAAATTTGTCAATGACAGAGCAGAAAGCGAAGGTGAAGTATTTGAC 1475  
Db 182 TGTATCGGAAAGTTTGTCAATGACAGAGCAGAAAGCGAAGGTGAAGTATTTGAC 241  
Qy 1476 GTGACTGTGCCCCCTGACAGTCCGTGTGAAGGACTCTGAAGTCAATCTCAGGAGGCTTCC 1535  
Db 242 GTGACTGTGCCCCCTGACAGTCCGTGTGAAGGACTCTGAAGTCAATCTCAGGAGGCTTCC 301  
Qy 1536 AAGCTGGGCTGCTGCTTTTATGACCCGCTGTGGGAGAAAGAAAGTGAAGT 1595  
Db 302 AAGCTGGGCTGCTGCTTTTATGACCCGCTGTGGGAGAAAGAAAGTGAAGT 361  
Qy 1596 CTCTATCATGTTCCGGGGCGTCTGTCATCAGGTGATGCTGTCGACAGTGAAGCCCTCCG 1655  
Db 362 CTCTATCATGTTCCGGGGCGTCTGTCATCAGGTGATGCTGTCGACAGTGAAGCCCTCCG 421  
Qy 1656 ATACCAAGACGTCCACAGATGATACAGATGAATAACTGCCAAGAACAGATTTT 1715  
Db 422 ATACCAAGACGTCCACAGATGATACAGATGAATAACTGCCAAGAACAGATTTT 481  
Qy 1716 AAAAGCCGCAAAAATCTTTTCTGGAGCTCAAAATTTGAAATGAAATCCAGA 1775  
Db 482 AAAAGCCGCAAAAATCTTTTCTGGAGCTCAAAATTTGAAATGAAATCCAGA 541  
Qy 1776 CATCAGATG-TTTTATTTATATATATATATATGAAGGTGATCAATTAATAT 1834  
Db 542 CATCAGATGTTTATTTATATATATATATATATGAAGGTGATCAATTAATAT 601  
Qy 1835 GAAGGACATGACACACCCAGTCTTTGAGGGTCTGGGGGTGAGACTGAGGACGCC 1894  
Db 602 GAAGGACATGACACACCCAGTCTTTGAGGGTCTGGGGGTGAGACTGAGGACGCC 661  
Qy 1895 ACTGGAACACAGATGACAGCTGGCCCATGAGCTGTTTCCAAAGATCAAGTCTGAGG 1954  
Db 662 ACTGGAACACAGATGACAGCTGGCCCATGAGCTGTTTCCAAAGATCAAGTCTGAGG 721  
Qy 1955 GAAGGAGCTGAGCCCTGACTCCGCTGTGTCCGAGCAGAGTGTGACCCGAGCCCG 2014  
Db 722 GAAGGAGCTGAGCCCTGACTCCGCTGTGTCCGAGCAGAGTGTGACCCGAGCCCG 781  
Qy 2015 CCTGTAGTCTTGGCTGGGCTGTGAGAGTGTCTGTGAGACACCTGCTTACCAACGA 2074  
Db 782 CCTGTAGTCTTGGCTGGGCTGTGAGAGTGTCTGTGAGACACCTGCTTACCAACGA 841  
Qy 2075 GCGTAGACCACTTCTGAGTCCACGCTGAACATGGGAAACAACCTGAAAGAGGCGAGG 2134  
Db 842 GCGTAGACCACTTCTGAGTCCACGCTGAACATGGGAAACAACCTGAAAGAGGCGAGG 901  
Qy 2135 CTCCCGGTCAAGGAGCCTGTGCTGTGCTGCTTCCATGACACACTTCTCTGTGAAT 2194  
Db 902 CTCCCGGTCAAGGAGCCTGTGCTGTGCTGCTTCCATGACACACTTCTCTGTGAAT 961  
Qy 2195 ATTACTGCTGAATCTGAGACAGATTGCGGGTTTATATAAAGTCTTTTATCTGAAGA 2254  
Db 962 ATTACTGCTGAATCTGAGACAGATTGCGGGTTTATATAAAGTCTTTTATCTGAAGA 1021  
Qy 2255 AAGCGGTTGGAAATTAATGCTGCTTTTCCCACTCCGAGAGTGTCAAGTATTTCA 2314  
Db 1022 AAGCGGTTGGAAATTAATGCTGCTTTTCCCACTCCGAGAGTGTCAAGTATTTCA 1081  
Qy 2315 CCGGCCCCCTCGCTTGGGACAGGGTATGTAATCTCCGATTCACAGGGGCTTACGCTTAC 2374  
Db 1082 CCGGCCCCCTCGCTTGGGACAGGGTATGTAATCTCCGATTCACAGGGGCTTACGCTTAC 1141  
Qy 2375 ACAGGTGCTTCCGCTATCCCGGTGGAAACGCTTCCGACACAGCGGGCTTGAAGTGGC 2434  
Db 1142 ACAGGTGCTTCCGCTATCCCGGTGGAAACGCTTCCGACACAGCGGGCTTGAAGTGGC 1201  
Qy 2435 CTGTGCTTCCGCTACCGCTGACACACCTCAAGTGTCAAGTGTGGCAAGGGGAGC 2494

```

Db      1202 CTGTGTCCCCCAACCCGCTGACCAACCCACTCCAGATGCTGCTGGCAAGGAGC 1261
Qy      2495 TCAAGAGAGAGAGACAGGCGCTTGGAGACATCAGACACACCAACCCAAAGGCGTGG 2554
Db      1262 TCAAGAGAGAGAGACAGGCGCTTGGAGACATCAGACACACCAACCCAAAGGCGTGG 1321
Qy      2555 ACCCCAGGCGCGGCGCTGATGCCAGAGATGAGCACTGCACTCCCGCTCTCTGACAGT 2614
Db      1322 ACCCCAGGCGCGGCGCTGATGCCAGAGATGAGCACTGCACTCCCGCTCTCTGACAGT 1381
Qy      2615 CCAAGGCTCTCAGAGAGAACACAGGCGCTGCTCCGAGACCTTCTCTGACAGCCCTTCTCT 2674
Db      1382 CCAAGGCTCTCAGAGAGAACACAGGCGCTGCTCCGAGACCTTCTCTGACAGCCCTTCTCT 1441
Qy      2675 CCAAGGCTCTCAGAGAGAACACAGGCGCTGCTCCGAGACCTTCTCTGACAGCCCTTCTCT 2734
Db      1442 CCAAGGCTCTCAGAGAGAACACAGGCGCTGCTCCGAGACCTTCTCTGACAGCCCTTCTCT 1501
Qy      2735 CTGGGCTGCAAGTAAGTTACGTGAGGCTCTCTCTCGGGGCTTGGAAAGTGGCAGCCATCA 2794
Db      1502 CTGGGCTGCAAGTAAGTTACGTGAGGCTCTCTCTCGGGGCTTGGAAAGTGGCAGCCATCA 1561
Qy      2795 GTTGTCTTGTGTGACCCCTCGAGAGAGAGCCCGCAGAGGTGTGTGCTGAGACAGCTGGCG 2854
Db      1562 GTTGTCTTGTGTGACCCCTCGAGAGAGAGCCCGCAGAGGTGTGTGCTGAGACAGCTGGCG 1621
Qy      2855 CGGGGGGCGCCCAAGCTGCGCGGCTTCAGCCCAACCAAGCTGTGTGCTGAGAGTTCAGAGCC 2914
Db      1622 CGGGGGGCGCCCAAGCTGCGCGGCTTCAGCCCAACCAAGCTGTGTGCTGAGAGTTCAGAGCC 1681
Qy      2915 AACCTCCCAAGCACTGTGTATCTGAGTAAAGGCTTAAAGAACTCTTCTCTGTGTTTGA 2974
Db      1682 AACCTCCCAAGCACTGTGTATCTGAGTAAAGGCTTAAAGAACTCTTCTCTGTGTTTGA 1741
Qy      2975 AGGAGTTCGGGTTGTCCAAATTCGTAAATTCATTCATTTTAAAGGTTTCTCT 3034
Db      1742 AGGAGTTCGGGTTGTCCAAATTCGTAAATTCATTCATTTT-AAAAAGGTTTCTCT 1800
Qy      3035 GACGGCCCAAGGCGCGGCTGAGAGGCTGTGTGATGAGCTGTGGGCGCCCGGCT 3094
Db      1801 GACGGCCCAAGGCGCGGCTGAGAGGCTGTGTGATGAGCTGTGGGCGCCCGGCT 1860
Qy      3095 TCCCGTGGCGCTTGTCCCGCAGGTGCTTGTGGGCAACCATCTCTGCGTTTCATTTGACGT 3154
Db      1861 TCCCGTGGCGCTTGTCCCGCAGGTGCTTGTGGGCAACCATCTCTGCGTTTCATTTGACGT 1920
Qy      3155 CGACTGTACAGAAAGGCACTCACAACAATAAACCCTTTCGAAAGCAGAAAAA 3213
Db      1921 CGACTGTACAGAAAGGCACTCACAACAATAAACCCTTTCGAAAGCAGAAAAA 1979

RESULT 7
LOCUS   BC014145          2125 bp    mRNA    linear    PRI 19-DEC-2003
DEFINITION Homo sapiens cDNA clone MGC:20641 IMAGE:4762069, complete cds.
ACCESSION BC014145.2 GI:40226303
VERSION   BC014145.2
KEYWORDS
SOURCE    Homo sapiens (human)
ORGANISM  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS   Klausner,R.D., Collins,F.S., Wagner,L.H., Grouse,L.H., Derge,J.G.,
          Strausberg,R.L., Feingold,E.A., Grouse,L.H., Schenck,C.M., Schler,G.D.,
          Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
          Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
          Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
          Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
          Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Teshitvuk,S.,
          Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
          Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
          McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,

```

```

TITLE     JOURNAL
REFERENCE 2 (bases 1 to 2125)
AUTHORS   Strausberg,R.
JOURNAL   Direct Submission
Submitted (10-SEP-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT   On Dec 19, 2003 this sequence version replaced gi:15559567.
Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland:
Web site: http://www.nisc.nih.gov/
Contact: nisc_mgc@nigr.nih.gov
Akhter,N., Aylee,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Lalic,P., Legaspi,R.,
Maduro,O.L., Masello,C., Maskeri,B., Mastrian,S.D., McCloskey,J.C.,
McDowell,J., Pearson,R., Scantirip,S., Thomas,P.J., Touchman,J.W.,
Tsurgoun,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
Young,A., Zhang,L.-H. and Green,E.D.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov
Series: IRAL Plate: 30 Row: 0 Column: 23
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 8922628.
Location/Qualifiers
1..2125
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:20641 IMAGE:4762069"
/issue_type="Brain, neuroblastoma"
/clone_lib="NIH MGC_47"
/lab_host="DH10B-R"
/notes="vector: pOTB7"
15..1538
/codon_start=1
/product="Unknown (protein for MGC:20641)"
/protein_id="AAH14145.1"
/db_xref="GI:15559568"
/translation="MATAISEEELDNEEDYSLNVRREASSELKAATYRLCMLYHPD
KRDDELKSQAERLNLVHAQAEVLSDDQTRAIYIYGRGLGEMGEVEVERRRPAB
IREEERLQREBERRLQDRTNPKCTIVGVADNLPFRYDEYDVSGSSPPOEIN
KMHISQSEAPLATDTAIIISGSLSTONGGGSINPALRYVTSAGMGELEFAGDL
OSAMTGLKLFRLITPRCVTTNCALQFSRGRPGLTTLVLANLKNKYGVLOFMNGI
OGAMTSLVDRKTSHTFVALQDGIPIHSPALISYHKKODDQTRVYKSLKAGPFGTV
VAYGKERISRSVLAIVSVGPGVSLKVKLEKQRESAATDVLOKKQKESAVRL
MQESVARIIEAESRMGLIIVAWGKGVNDSRSEKVKYIVDTVPLOCIVKDXLI
ILEAKAGIPGVDPDCEVEERKILKLYQFRGLVGVVLDSEALIPQOSHRIIDTG"
54..260
/notes="nafj. Region: DnaJ domain. DnaJ domains (J-domains)
are associated with hsp70 heat-shock system and it is
thought that this domain mediates the interaction.
DnaJ-domain is therefore part of a chaperone (protein

```

misc\_feature

folding) system. The T-antigens are confirmed as DnaU  
containing domains from literature"  
/db\_xref="CDD:pfam00226"

## ORIGIN

Query Match 60.2%; Score 1934; DB 9; Length 2125;  
Best Local Similarity 93.1%; Pred. No. 0;  
Matches 2100; Conservative 0; Mismatches 0; Indels 156; Gaps 1;

QY 1 GAAAGGTTGCGAAGATGGCGACGGCTTTGAGCGAGAGAGCTGGACATTAAGACTATT 60  
DB 1 GAAAGGTTGCGAAGATGGCGACGGCTTTGAGCGAGAGAGCTGGACATTAAGACTATT 60  
QY 61 ACTGTTGCTGAAGAGTGGCGAGGGAGGCTCTTCTGAAGAGCTGAAGAGCTGCTTACGGGA 120  
DB 61 ACTGTTGCTGAAGAGTGGCGAGGGAGGCTCTTCTGAAGAGCTGAAGAGCTGCTTACGGGA 120  
QY 121 GGCTCTGATGCTCTACCATCCAGACAGACAGAGCCAGAGCTCAAGTCAAGGCGG 180  
DB 121 GGCTCTGATGCTCTACCATCCAGACAGACAGAGCCAGAGCTCAAGTCAAGGCGG 180  
QY 181 AAGCACTGTTTAACTTGTTCACAGAGGCTTATGAAGTGTGACCCCAAAACAGGG 240  
DB 181 AAGCACTGTTTAACTTGTTCACAGAGGCTTATGAAGTGTGACCCCAAAACAGGG 240  
QY 241 CCATCTATGATATATATGAGAGAGAGGACTGSAATGSAAGATGGAGGTTGTGAAA 300  
DB 241 CCATCTATGATATATATGAGAGAGAGGACTGSAATGSAAGATGGAGGTTGTGAAA 300  
QY 301 GAGAGAGAACCCCTGCTGTAATTCGAGAGAGATTGAGCGGCTGACAGAGAGAGAG 360  
DB 301 GAGAGAGAACCCCTGCTGTAATTCGAGAGAGATTGAGCGGCTGACAGAGAGAGAG 360  
QY 361 AGAGAGATTGCGAGCGGAACCAATCCAAAGGAGATAGCGTTGAGTATGATGCCA 420  
DB 361 AGAGAGATTGCGAGCGGAACCAATCCAAAGGAGATAGCGTTGAGTATGATGCCA 420  
QY 421 CCGACCTTTTGTATGCGTATATGAGAGATGSAAGATGTCGGGCAAGTGTTCGGC 480  
DB 421 CCGACCTTTTGTATGCGTATATGAGAGATGSAAGATGTCGGGCAAGTGTTCGGC 480  
QY 481 AGATTGAATTAATTAATGACATATCCAGTCCATTGAGGCAACCTTGAACAGCAG 540  
DB 481 AGATTGAATTAATTAATGACATATCCAGTCCATTGAGGCAACCTTGAACAGCAG 540  
QY 541 ACAAGCCATCTCTTGTGAAGCTCTCAACCCAGATGAATGAATGAGAGGTTCCATT 600  
DB 541 ACAAGCCATCTCTTGTGAAGCTCTCAACCCAGATGAATGAATGAGAGGTTCCATT 600  
QY 601 ACTTGGGCTGACGAGTAACCTTGGGAAGGATGGGAGATGGAAATTTGGAAGTGG 660  
DB 601 ACTTGGGCTGACGAGTAACCTTGGGAAGGATGGGAGATGGAAATTTGGAAGTGG 660  
QY 661 GAGACTTAAGGGGCTTTGTTGGTCTCAAGTGTTCCTGATCTCAACCAAGATGCT 720  
DB 661 GAGACTTAAGGGGCTTTGTTGGTCTCAAGTGTTCCTGATCTCAACCAAGATGCT 720  
QY 721 TTGTGACAAACAACTGTGCTGTGAGTTTCAATCCGTAATCGAACCCGGCTGACA 780  
DB 721 TTGTGACAAACAACTGTGCTGTGAGTTTCAATCCGTAATCGAACCCGGCTGACA 780  
QY 781 CTGTCTTAAGTCTGGAACCTTAACAAGAACACCGTGGGCTACCTGAGTGGAGTGG 840  
DB 781 CTGTCTTAAGTCTGGAACCTTAACAAGAACACCGTGGGCTACCTGAGTGGAGTGG 840  
QY 841 TCCAGTGAAGCATTAAGCATGCTCCGAGACACTTAACCAAGCACTTCACTGGG 900  
DB 841 TCCAGTGAAGCATTAAGCATGCTCCGAGACACTTAACCAAGCACTTCACTGGG 900  
QY 901 CCTGCAAGCTGGGAATCCCTCACTCTTTGCACTGATCAGTATCAGCAAAATTCAG 960  
DB 901 CCTGCAAGCTGGGAATCCCTCACTCTTTGCACTGATCAGTATCAGCAAAATTCAG 960

QY 961 ATGACATCAGACTCGTGTGAAGAGATCCCTCAAGCAGGCTTCTTTGGACGCTGGTGG 1020  
DB 961 ATGACATCAGACTCGTGTGAAGAGATCCCTCAAGCAGGCTTCTTTGGACGCTGGTGG 1020  
QY 1021 AGTACGAGCTGAGAGAGATCTCCAGGACAGCGCTTTGGTGCAGCTGTCAAGCGTTG 1080  
DB 1021 AGTACGAGCTGAGAGAGATCTCCAGGACAGCGCTTTGGTGCAGCTGTCAAGCGTTG 1080  
QY 1081 GAGTTCACAGGCGCTTCTCTCAAGTCAAGCTCAACAGGGCCAGTCAACATCTTCT 1140  
DB 1081 GAGTTCACAGGCGCTTCTCTCAAGTCAAGCTCAACAGGGCCAGTCAACATCTTCT 1140  
QY 1141 TCCCTATTCACTTGAACGAGCTTCTGACCAGCGCATGTTCTATGCCAGCGTGGGCG 1200  
DB 1141 TCCCTATTCACTTGAACGAGCTTCTGACCAGCGCATGTTCTATGCCAGCGTGGGCG 1200  
QY 1201 CTCTAGTGTCTACTTTCATGCAACCGTCTGATCAACCAATACCTGAGGCTGAGA 1260  
DB 1201 CTCTAGTGTCTACTTTCATGCAACCGTCTGATCAACCAATACCTGAGGCTGAGA 1260  
QY 1261 AAGAGAAATTTGAGAGACAGAGAGAAAGCCCGCACCGATGTGCTGACAGAGAGC 1320  
DB 1261 AAGAGAAATTTGAGAGACAGAGAGAAAGCCCGCACCGATGTGCTGACAGAGAGC 1320  
QY 1310 -----AAGAAATTTGAGAGACAGAGAGAAAGCCCGCACCGATGTGCTGACAGAGAGC 1364  
DB 1310 -----AAGAAATTTGAGAGACAGAGAGAAAGCCCGCACCGATGTGCTGACAGAGAGC 1364  
QY 1321 AAGAGCGGAGTCCGCTGTCCGCTGATGCAAGATCTGTCGGAAGATTAATTGAGGCG 1380  
DB 1321 AAGAGCGGAGTCCGCTGTCCGCTGATGCAAGATCTGTCGGAAGATTAATTGAGGCG 1380  
QY 1365 AAGAGCGGAGTCCGCTGTCCGCTGATGCAAGATCTGTCGGAAGATTAATTGAGGCG 1424  
DB 1365 AAGAGCGGAGTCCGCTGTCCGCTGATGCAAGATCTGTCGGAAGATTAATTGAGGCG 1424  
QY 1381 AAGAGTCCAGAAATGGGCTCATCATGCTGATGCTGATGCTGATGCTGATGCTGATGCA 1440  
DB 1381 AAGAGTCCAGAAATGGGCTCATCATGCTGATGCTGATGCTGATGCTGATGCTGATGCA 1440  
QY 1425 AAGAGTCCAGAAATGGGCTCATCATGCTGATGCTGATGCTGATGCTGATGCTGATGCA 1484  
DB 1425 AAGAGTCCAGAAATGGGCTCATCATGCTGATGCTGATGCTGATGCTGATGCTGATGCA 1484  
QY 1441 AAGAGCGGAGTCCGCTGTCCGCTGATGCAAGATCTGTCGGAAGATTAATTGAGGCG 1500  
DB 1441 AAGAGCGGAGTCCGCTGTCCGCTGATGCAAGATCTGTCGGAAGATTAATTGAGGCG 1500  
QY 1485 AAGAGCGGAGTCCGCTGTCCGCTGATGCAAGATCTGTCGGAAGATTAATTGAGGCG 1544  
DB 1485 AAGAGCGGAGTCCGCTGTCCGCTGATGCAAGATCTGTCGGAAGATTAATTGAGGCG 1544  
QY 1501 TGAAGGACTGAGAGTCACTCAACGAGGCTCCAAAGGCTGGGCTGCTGCTTTATG 1560  
DB 1501 TGAAGGACTGAGAGTCACTCAACGAGGCTCCAAAGGCTGGGCTGCTGCTTTATG 1560  
QY 1545 TGAAGGACTGAGAGTCACTCAACGAGGCTCCAAAGGCTGGGCTGCTGCTTTATG 1604  
DB 1545 TGAAGGACTGAGAGTCACTCAACGAGGCTCCAAAGGCTGGGCTGCTGCTTTATG 1604  
QY 1561 ACCGCTGTGCGGAGAGAGAGAACTGAAAGTGTCTATCAGTTCGCGGCGCTCTGCG 1620  
DB 1561 ACCGCTGTGCGGAGAGAGAGAACTGAAAGTGTCTATCAGTTCGCGGCGCTCTGCG 1620  
QY 1605 ACCGCTGTGCGGAGAGAGAGAACTGAAAGTGTCTATCAGTTCGCGGCGCTCTGCG 1664  
DB 1605 ACCGCTGTGCGGAGAGAGAGAACTGAAAGTGTCTATCAGTTCGCGGCGCTCTGCG 1664  
QY 1621 ATCAGGATGATGCTGCTGAGACATGAGGCTCTCGGATACCAAGCAGTCCCAAGATG 1680  
DB 1621 ATCAGGATGATGCTGCTGAGACATGAGGCTCTCGGATACCAAGCAGTCCCAAGATG 1680  
QY 1665 ATCAGGATGATGCTGCTGAGACATGAGGCTCTCGGATACCAAGCAGTCCCAAGATG 1724  
DB 1665 ATCAGGATGATGCTGCTGAGACATGAGGCTCTCGGATACCAAGCAGTCCCAAGATG 1724  
QY 1681 ATCAGATGATTAACCTGCAAGAACCAAGATTTTAAAGGCGCAAAAATCTTTTCT 1740  
DB 1681 ATCAGATGATTAACCTGCAAGAACCAAGATTTTAAAGGCGCAAAAATCTTTTCT 1740  
QY 1725 ATCAGATGATTAACCTGCAAGAACCAAGATTTTAAAGGCGCAAAAATCTTTTCT 1784  
DB 1725 ATCAGATGATTAACCTGCAAGAACCAAGATTTTAAAGGCGCAAAAATCTTTTCT 1784  
QY 1741 GGGAGTCTAACAATTTGGAATGAAGAAACCCAGACATCAGATGTTTATTTATATTA 1800  
DB 1741 GGGAGTCTAACAATTTGGAATGAAGAAACCCAGACATCAGATGTTTATTTATATTA 1800  
QY 1785 GGGAGTCTAACAATTTGGAATGAAGAAACCCAGACATCAGATGTTTATTTATATTA 1844  
DB 1785 GGGAGTCTAACAATTTGGAATGAAGAAACCCAGACATCAGATGTTTATTTATATTA 1844  
QY 1801 TTATTATGAAGGTTGATCAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1860  
DB 1801 TTATTATGAAGGTTGATCAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1860  
QY 1861 TTGAGGATGCTGGGAGTGAAGACTGAGGACCCCACTGGGAAACCAAGCTCAGCTGGCC 1920  
DB 1861 TTGAGGATGCTGGGAGTGAAGACTGAGGACCCCACTGGGAAACCAAGCTCAGCTGGCC 1920  
QY 1905 TTGAGGATGCTGGGAGTGAAGACTGAGGACCCCACTGGGAAACCAAGCTCAGCTGGCC 1964  
DB 1905 TTGAGGATGCTGGGAGTGAAGACTGAGGACCCCACTGGGAAACCAAGCTCAGCTGGCC 1964  
QY 1921 CATGAGCTGTTTCCCAAGATCAGTTCCTGAGAGAAAGGCTCTGAGCTGAGCTGCGCTG 1980  
DB 1921 CATGAGCTGTTTCCCAAGATCAGTTCCTGAGAGAAAGGCTCTGAGCTGAGCTGCGCTG 1980  
QY 1965 CATGAGCTGTTTCCCAAGATCAGTTCCTGAGAGAAAGGCTCTGAGCTGAGCTGCGCTG 2040  
DB 1965 CATGAGCTGTTTCCCAAGATCAGTTCCTGAGAGAAAGGCTCTGAGCTGAGCTGCGCTG 2040  
QY 1981 TGTCCGAGCAACAGTGTGACCGAGCGCGCTGTAATTTGAGTGTGAGTGTGAGTGTG 2040  
DB 1981 TGTCCGAGCAACAGTGTGACCGAGCGCGCTGTAATTTGAGTGTGAGTGTGAGTGTG 2040  
QY 1825 TGTCCGAGCAACAGTGTGACCGAGCGCGCTGTAATTTGAGTGTGAGTGTGAGTGTG 1884  
DB 1825 TGTCCGAGCAACAGTGTGACCGAGCGCGCTGTAATTTGAGTGTGAGTGTGAGTGTG 1884  
QY 2041 GGTGTGTGTGAACACCTGCTCCCTCAACCAAGAGGTTGAGCACTTCTGCAATCCAGCG 2100  
DB 2041 GGTGTGTGTGAACACCTGCTCCCTCAACCAAGAGGTTGAGCACTTCTGCAATCCAGCG 2100

Db 1885 GGTGTCTGTGAGACACCTCCCTCACACAGAGAGGTGAGCCACTTCTCAGTCCACGC 1944  
Qy 2101 TGAACATGGGAACAACCTGAAAGAGAGAGGCTCCCGGTGAGGAGAGCTCTGCTG 2160  
Db 1945 TGAACATGGGAACAACCTGAAAGAGAGAGGCTCCCGGTGAGGAGAGCTCTGCTG 2004  
Qy 2161 CTGGCTCCCATGACCACTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2220  
Db 2005 CTGGCTCCCATGACCACTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2064  
Qy 2221 GCGGGTTATTAACCTGCTTTTATCTGAGACAA 2256  
Db 2065 GCGGGTTATTAACCTGCTTTTATCTGAGACAA 2100  
RESULT 8  
AX748271 2454 bp mRNA linear PAT 20-JUN-2003  
LOCUS AX748271 Sequence 1796 from Patent EP1308459.  
DEFINITION AX748271  
ACCESSION AX748271 GI:32132659  
VERSION AX748271.1 GI:32132659  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;  
Eukaryota; Metazoa; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1  
Izogai,T., Sugiyama,T., Otsuki,T., Wakamatsu,A., Sato,H., Ishii,S.,  
Yamamoto,J.I., Isono,Y., Hio,Y., Otsuka,K., Nagai,K., Irie,R.,  
Tamechika,I., Seki,N., Yoshikawa,T., Otsuka,M., Nagahara,K. and  
Masuko,Y.  
Full-length cDNA sequences  
Patent: EP 1308459-A 1796 07-MAY-2003:  
Helix Research Institute (JP) ; Research Association for  
Biotechnology (JP)  
LOCATION/Qualifiers  
FEATURES  
source 1..2454  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
ORIGIN  
Query Match 57.6%; Score 1851.6; DB 6; Length 2454;  
Best Local Similarity 99.7%; Pred. No. 0;  
Matches 1865; Conservative 0; Mismatches 4; Indels 1; Gaps 1;  
Qy 1338 GTCCGGCTGATGCAAGATCTGTCCGAAAGATATTGAGGAGAGAGTCCAGATGGGC 1397  
Db 585 GTCCGGCTGATGCAAGATCTGTCCGAAAGATATTGAGGAGAGTCCAGATGGGC 644  
Qy 1398 CTATCATGCTCAATGCTGTGTAAGGAGTTTGTCAATGCAAGAGCAGAGAGAGCGAG 1457  
Db 645 CTATCATGCTCAATGCTGTGTAAGGAGTTTGTCAATGCAAGAGCAGAGAGAGCGAG 704  
Qy 1458 AAGGTGAAGTGAATTGCTGATGCTGCTCCCTGCAAGTCTGCTGCTGCTGCTGCTG 1517  
Db 705 AAGGTGAAGTGAATTGCTGATGCTGCTCCCTGCAAGTCTGCTGCTGCTGCTGCTG 764  
Qy 1518 ATCTTCAGGAGGCTCTCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1577  
Db 765 ATCTTCAGGAGGCTCTCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 824  
Qy 1578 GAGAGAGAGCTGAAGTGTCTATCAGTTCCGGGGGCTCTGATCAGAGTGAAGTGTGCTG 1637  
Db 825 GAGAGAGAGCTGAAGTGTCTATCAGTTCCGGGGGCTCTGATCAGAGTGAAGTGTGCTG 884  
Qy 1638 GACAGTGAAGGCTCTCCGATACCAAGACAGTCCACAGATGATACAGATGATGATAACT 1697  
Db 885 GACAGTGAAGGCTCTCCGATACCAAGACAGTCCACAGATGATACAGATGATGATAACT 944  
Qy 1698 GCCAAGAACCAAGATTTTAAAGGCGCAAAAATCTTTTCTGGAGTCTACAAATTTG 1757

Db 945 GCCAAGAACCAAGATTTTAAAGGCGCAAAAATCTTTTCTGGAGTCTACAAATTTG 1004  
Qy 1758 GAAATGAAAAAACCAGACATGATGTTTTATTTATTTATTTATTTATTAAGGTGT 1817  
Db 1005 GAAATGAAAAAACCAGACATGATGTTTTATTTATTTATTTATTTATTAAGGTGT 1064  
Qy 1818 ACCATTATCAATTATGTAAGGAGATGACAGACACCCAGCTTTTGAGGCTGCTGGGGT 1877  
Db 1065 ACCATTATCAATTATGTAAGGAGATGACAGACACCCAGCTTTTGAGGCTGCTGGGGT 1124  
Qy 1878 AGGACTGAGGAGCCCACTGGGAACAGACTGCAAGCTGCGCCATGCTGTTTTCCAA 1937  
Db 1125 AGGACTGAGGAGCCCACTGGGAACAGACTGCAAGCTGCGCCATGCTGTTTTCCAA 1184  
Qy 1938 GGATCAGTTCCTGAGGGAAGGCTCTGGCCCTGCACTCCGCTGTGTCCGAGACACG 1997  
Db 1185 GGATCAGTTCCTGAGGGAAGGCTCTGGCCCTGCACTCCGCTGTGTCCGAGACACG 1244  
Qy 1998 CTGACCGAGCCCGCCGCTGATGTTCTTGCTGGGTCTGAGGTGTCTGAGACACC 2057  
Db 1245 CTGACCGAGCCCGCCGCTGATGTTCTTGCTGGGTCTGAGGTGTCTGAGACACC 1304  
Qy 2058 CTGGCTTCAACACAGAGAGCTGAGCCATTCTGCACTGCAAGCTGAAACATGGGAAACAC 2117  
Db 1305 CTGGCTTCAACACAGAGAGCTGAGCCATTCTGCACTGCAAGCTGAAACATGGGAAACAC 1364  
Qy 2118 CTGAAAGAGAGAGGCTCCCGGTGAGGAGACCTGCTGCTGCTGCTGCTGCTGCTGCTG 2177  
Db 1365 CTGAAAGAGAGAGGCTCCCGGTGAGGAGACCTGCTGCTGCTGCTGCTGCTGCTGCTG 1424  
Qy 2178 CCTCTCTGCTGAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2237  
Db 1425 CCTCTCTGCTGAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1484  
Qy 2238 CTTTTTATCTGGAACAAACGCGTTTGAAATTAATTAATTAATTAATTAATTAATTA 2297  
Db 1485 CTTTTTATCTGGAACAAACGCGTTTGAAATTAATTAATTAATTAATTAATTAATTA 1544  
Qy 2298 CTGCTCAATCAATTCACACGCGCCCTCGGCTTGGAACAGGTAAGTAACTCCGATCC 2357  
Db 1545 CTGCTCAATCAATTCACACGCGCCCTCGGCTTGGAACAGGTAAGTAACTCCGATCC 1604  
Qy 2358 CAGGCTTACGCTGACACAGGTGCTTCCGATCCCGGTGGAACAGCCCTGACAC 2417  
Db 1605 CAGGCTTACGCTGACACAGGTGCTTCCGATCCCGGTGGAACAGCCCTGACAC 1664  
Qy 2418 AGCGGCTTGAAGTGGCTGTGTCTCTCAAGCTGCAACCTCCAGATGCTGAG 2477  
Db 1665 AGCGGCTTGAAGTGGCTGTGTCTCTCAAGCTGCAACCTCCAGATGCTGAG 1724  
Qy 2478 TGCTGGGCAAGGAGCTCAAGAGGACAGGACAGGAGCTTGCAAGACATGACACAC 2537  
Db 1725 TGCTGGGCAAGGAGCTCAAGAGGAGGACAGGAGCTTGCAAGACATGACACAC 1784  
Qy 2538 CCAACCCAAAGGCTGAGACCCAGGCGCGCTGTGTAACCAAGAGTGGACCTGACG 2597  
Db 1785 CCAACCCAAAGGCTGAGACCCAGGCGCGCTGTGTAACCAAGAGTGGACCTGACG 1844  
Qy 2598 TCCCGCTCTGCAAGTCCAGGCTCTCAAGAGACCAAGGCTGTGCTCCGAGCT 2657  
Db 1845 TCCCGCTCTGCAAGTCCAGGCTCTCAAGAGACCAAGGCTGTGCTCCGAGCT 1904  
Qy 2658 TCCCTCAGACCTCTCTCCAGTGCACCTTGGAATGAGATGAGAGGAGCTAGAGAC 2717  
Db 1905 TCCCTCAGACCTCTCTCCAGTGCACCTTGGAATGAGATGAGAGGAGCTAGAGAC 1964  
Qy 2718 CCTTCAGGCTGAGACCTGAGCTGAGCTGAGTAAATTAAGTGAAGCTGTCTCGGGCT 2777  
Db 1965 CCTTCAGGCTGAGACCTGAGCTGAGCTGAGTAAATTAAGTGAAGCTGTCTCGGGCT 2024  
Qy 2778 GGAAGTGGAGCATCATGTTGCTTTGCTGACCCCTCGAGAGAGGCTCGACAGTGT 2837  
Db 2025 GGAAGTGGAGCATCATGTTGCTTTGCTGACCCCTCGAGAGAGGCTCGACAGTGT 2084

QY	2838	GGCTGAGCAGCTGGCGGGGGGGCCCCAAGCTGGCGCGGCTCCAGGCCA	CCCAAGCT	2897
Db	2085	GGCTGAGCAGCTGGCGGGGGGGCCCCAAGCTGGCGCGGCTCCAGGCCA	CCCAAGCT	2144
QY	2898	GTTGCTGAGTCAAGCCAACTCCCAAGCACTGATATCTAGTAA	CGGCTAAGAACTCC	2957
Db	2145	GTTGCTGAGTCAAGCCAACTCCCAAGCACTGATATCTAGTAA	CGGCTAAGAACTCC	2204
QY	2958	TTCTCTGGTTTTGAAAAGCAGTTGGGTTGTCCAAATCTGTAA	CATTCATCTCATTTT	3017
Db	2205	TTCTCTGGCTTTGAAAAGCAGTTGGGTTGTCCAAATCTGTAA	CATTCATCTCATTTT	2264
QY	3018	TTAAAAAGTTTTCTCTGACGGGCCCAAGGCCCAAGCGCGGTGAGC	TCGTGTTTGCATG	3077
Db	2265	TT-AAAAAGTTTTCTCTGACGGGCCCAAGGCCCAAGCGCGGTGAGC	TCGTGTTTGCATG	2323
QY	3078	AGCCTGGGCCCCGGGCTTCCCGTGCGCTCTGCGCAGAGTCTTT	CTGGGACCACTCTC	3137
Db	2234	AGCCTGGGCCCCGGGCTTCCCGTGCGCTCTGCGCAGAGTCTTT	CTGGGACCACTCTC	2383
QY	3138	TGCGTTTCATTTGACAGTGCATGTACAGAAAGCACTCA	CAATAACTTTCTGAAA	3197
Db	2384	TGCGTTTCATTTGACAGTGCATGTACAGAAAGCACTCA	CAATAACTTTCTGAAA	2443
QY	3198	GCAGAAAAA	3207	
Db	2444	GCAGAAAAA	2453	

LOCUS	AK093709	2454 bp	mRNA	linear	PRI 06-SEP-2003
DEFINITION	Homo sapiens cDNA FLJ36390 fis, clone THYMJ2008990.				
DESCRIPTION					

VERSION AK093709.1 GI:21752634  
KEYWORDS oligo capring; fis (full insert sequence).  
SOURCE Homo sapiens (human)

ORGANISM

REFERENCE  
AUTHORS

Osshima, A., Takahashi-Fujii, A., Tanase, T., Imose, N., Takeuchi, K., Arita, M., Mitsuhashino, K., Yuuki, H., Hara, H., Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Ota, T., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Yamashita, H., Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K., Wagaratsuna, M., Murakawa, K., Kanehori, K., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Nagahata, K., Masuno, Y., Nagai, K., and Inagaki, T.

**TITLE** NEDO human cDNA sequencing project  
**JOURNAL** Unpublished

## REFERENCE

## AUTHORS

**JOURNAL**

**COMMENT**

Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.

## FEATURES

Source

```
/organism="Homo sapiens"  
/mol_type="mRNA"  
/db_xref="taxon:9606"  
/clone="THYMU2008990"  
/tissue_type="thymus"
```

```

/clone_lib="THYMU2"
/note="cloning vector: pME18SFL3"
ORIGIN

```

Query Match	57.6%	Score 1851.6;	DB 9;	Length 2454;
Best Local Similarity	99.7%;	Pred. No. 0;		
Matches 1865; Conservative	0;	Mismatches 4;	Indels 1;	Gaps 1;

QY	1338	GTCCGGCTGATGCAAGAAATCTGTCCGAAGATTAATTGAGCGAAGAAAGATCCAGAATGGCC	1397
Db	585	GTCCGGCTGATGCAAGAAATCTGTCCGAAGATTAATTGAGCGAAGAAAGATCCAGAATGGCC	644
QY	1398	CTCATCATCTGCAATGCCCTGCTGATCGGAAAGTTTGTCAATGACAAGAGCAGAAAGACGAG	1457
Db	645	CTCATCATCTGCAATGCCCTGCTGATCGGAAAGTTTGTCAATGACAAGAGCAGAAAGACGAG	704
QY	1458	AAGGTGAAGGTGATTTGACGTGACCTGTGCAGTGTCTGTGTGAAGAACTCCGAAGCTC	1511
Db	705	AAGGTGAAGGTGATTTGACGTGACCTGTGCAGTGTCTGTGTGAAGAACTCCGAAGCTC	764
QY	1518	ATCTCTACCGAGGCTCCAAAGGCTGGGCTGCTGGCTTTTATGACCCGCTGTGGGGAGAA	1577
Db	765	ATCTCTACCGAGGCTCCAAAGGCTGGGCTGCTGGCTTTTATGACCCGCTGTGGGGAGAA	824
QY	1578	GAGAAAGAACTGAAGAGTGTCTATCAGTTCGGGGGCGTCCCTGCATCAAGTATGGTGTCTG	1637
Db	825	GAGAAAGAACTGAAGAGTGTCTATCAGTTCGGGGGCGTCCCTGCATCAAGTATGGTGTCTG	884
QY	1638	GACAGTGAAGCCCTTCGGATTCGAAGACATGCCACAGATGCAATACAGATGATTAACCT	1697
Db	885	GACAGTGAAGCCCTTCGGATTCGAAGACATGCCACAGATGCAATACAGATGATTAACCT	944
QY	1698	GCCAAAGAACAGATTTTTTAAAAAGCCGCAAAAAATCTTTTCTGSGAGCTTACAAATTGG	1755
Db	945	GCCAAAGAACAGATTTTTTAAAAAGCCGCAAAAAATCTTTTCTGSGAGCTTACAAATTGG	1006
QY	1758	GAAATGAAAAAACCCAGACATCAGATGTTTTTATATATATTAATTAATTAAGAGTGT	1817
Db	1005	GAAATGAAAAAACCCAGACATCAGATGTTTTTATATATATTAATTAATTAAGAGTGT	1066
QY	1818	ACCATTTATCAATTATGTGAAGGGAACATGCAACACCCAGCTTTTGAAGGCTGTGGGGGT	1877
Db	1065	ACCATTTATCAATTATGTGAAGGGAACATGCAACACCCAGCTTTTGAAGGCTGTGGGGGT	1124
QY	1878	AGGACTGAGGAGGCGCCACTGGGAAACAGACTGAGGCTGGCCCATGCTGTTTTCCAA	1937
Db	1125	AGGACTGAGGAGGCGCCACTGGGAAACAGACTGAGGCTGGCCCATGCTGTTTTCCAA	1184
QY	1938	GGATCAGTTCCTGAGAGGGAAGGCTCTGGCCCTGACTCCGCTGTGTCCCGAGCACACGTG	1997
Db	1185	GGATCAGTTCCTGAGAGGGAAGGCTCTGGCCCTGACTCCGCTGTGTCCCGAGCACACGTG	1244
QY	1998	CTGACCGGAGCCGCGCCCTGTAGTTCTTGGCTGGGTCTGAGAGTGTCTGTGAGCACCC	2057
Db	1245	CTGACCGGAGCCGCGCCCTGTGTAGTTCTTGGCTGGGTCTGAGAGTGTCTGTGAGCACCC	1304
QY	2058	CTGGCCCTCAACCAAGAGAGCTGAGGCCACTTCTGAGTCCACGCTGAAACATGGGAAACAAAC	2117
Db	1305	CTGGCCCTCAACCAAGAGAGCTGAGGCCACTTCTGAGTCCACGCTGAAACATGGGAAACAAAC	1364
QY	2118	CTGAAAAAGCAGGCGAGGCTCCCGGTCAAGGAGCTCTGTGTGTGCTGTCCCAATGACCA	2177
Db	1365	CTGAAAAAGCAGGCGAGGCTCCCGGTCAAGGAGCTCTGTGTGTGCTGTCCCAATGACCA	1424
QY	2178	CTCTCTCTCTGTAATATATTACTGCTTGAATCTGAGCGAGTTCCGGGTTTTATTAACCTG	2237
Db	1425	CTCTCTCTCTGTAATATATTACTGCTTGAATCTGAGCGAGTTCCGGGTTTTATTAACCTG	1484
QY	2238	CTTTTATATCGAGAACAAAGGGGTTTGGAAATTAGTGTGTTTTTCCCACTCCACAGAG	2297
Db	1485	CTTTTATATCGAGAACAAAGGGGTTTGGAAATTAGTGTGTTTTTCCCACTCCACAGAG	1544
QY	2298	CTGTCTCAAGTATTCACCGGCGCCCTTCGGCTTGGGACAGGGTAGTGTAACTCCCGATCC	2357

Db 1545 CTCTCAAGTCATTCCACCGGCCCTCGGCTTGAGGACAGGTAAGTAATCCCATCC 1604  
 QY 2358 CAGGGGCTAGCCCTGACAGAGTGGCTTCCGTAATCCGGGTGGAAAAAGCCCTGACACC 2417  
 Db 1605 CAGGGGCTAGCCCTGACAGAGTGGCTTCCGTAATCCGGGTGGAAAAAGCCCTGACACC 1664  
 QY 2418 AGCGGGCTTAGAGTGGCTTGTGTCCTCAACCGCTGACACCACTTCAGAGTGCAG 2477  
 Db 1665 AGCGGGCTTAGAGTGGCTTGTGTCCTCAACCGCTGACACCACTTCAGAGTGCAG 1724  
 QY 2478 TGGTGGGCAAGGGCAGCTCAAGAGACAGACAGCGCTTGGAAGACATCAGACACAC 2537  
 Db 1725 TGGTGGGCAAGGGCAGCTCAAGAGACAGACAGCGCTTGGAAGACATCAGACACAC 1784  
 QY 2538 CCAACCCAAAGGGGTGAGACCCAGGCGCCGCTGTACCCAGAGTGGCACTGAGC 2597  
 Db 1785 CCAACCCAAAGGGGTGAGACCCAGGCGCCGCTGTACCCAGAGTGGCACTGAGC 1844  
 QY 2598 TCCCGGCTCTGAGAGTCCAGCGTCTCAAGAGAACACAGGGCTGTGCTCCGAGCCT 2657  
 Db 1845 TCCCGGCTCTGAGAGTCCAGCGTCTCAAGAGAACACAGGGCTGTGCTCCGAGCCT 1904  
 QY 2658 TCCCTTCAAGACCTTCTCTCCACGTCGCCACTTGGAGATGACAGCGAGCTAGGACC 2717  
 Db 1905 TCCCTTCAAGACCTTCTCTCCACGTCGCCACTTGGAGATGACAGCGAGCTAGGACC 1964  
 QY 2718 CCTTCAACGGGCTGAGACCTGGGCTGAGTAAGTTACGTAGAGCCCTGCTCGGGGCT 2777  
 Db 1965 CCTTCAACGGGCTGAGACCTGGGCTGAGTAAGTTACGTAGAGCCCTGCTCGGGGCT 2024  
 QY 2778 GGAAGTGGCAGCAGCATAGTTGCTTGTGTCGACCCCTCGAGACAGCGCCGACAGGTGT 2837  
 Db 2025 GGAAGTGGCAGCAGCATAGTTGCTTGTGTCGACCCCTCGAGACAGCGCCGACAGGTGT 2084  
 QY 2838 GGGCTGAGACAGCTGGCGCGGGGGGCCCAAGCTGGCGCGGCTTCAGGCCCAACCAAGCT 2897  
 Db 2085 GGGCTGAGACAGCTGGCGCGGGGGGCCCAAGCTGGCGCGGCTTCAGGCCCAACCAAGCT 2144  
 QY 2898 GTTGTGAAGTCAAGGCGCAACCTCCCGACGACTGGTAATCTGAGTAAGGCTAAGACCTCC 2957  
 Db 2145 GTTGTGAAGTCAAGGCGCAACCTCCCGACGACTGGTAATCTGAGTAAGGCTAAGACCTCC 2204  
 QY 2958 TTCTCTGCTTTTGAAGAAGAGTTCGGGTTGTCGAATCTGTAACATTCATCTCAATTT 3017  
 Db 2205 TTCTCTGCTTTTGAAGAAGAGTTCGGGTTGTCGAATCTGTAACATTCATCTCAATTT 2264  
 QY 3018 TTTAAAAAGGTTTCTGATGAGCCCAAGGCGCGGCTGAGAGCTGTGTCAGT 3077  
 Db 2265 TTTAAAAAGGTTTCTGATGAGCCCAAGGCGCGGCTGAGAGCTGTGTCAGT 2323  
 QY 3078 AGGCTGGGCGCCGGGGCTTCCGAGCGCTGTGCGGAGGTCTTGAGGACCAATCTTC 3137  
 Db 2324 AGGCTGGGCGCCGGGGCTTCCGAGCGCTGTGCGGAGGTCTTGAGGACCAATCTTC 2383  
 QY 3138 TGGCTTTCATTTGACGTGACGTGTAAGAGAGCACTCACCAATTAACCTTCTGAAA 3197  
 Db 2384 TGGCTTTCATTTGACGTGACGTGTAAGAGAGCACTCACCAATTAACCTTCTGAAA 2443  
 QY 3198 GCAGAAAAA 3207  
 Db 2444 GCAGAAAAA 2453

RESULT 10  
 AK095386 3012 bp mRNA linear PRI 15-JUL-2002  
 LOCUS Homo sapiens cDNA FLJ38067 f1s, clone CTONG2015345, weakly similar  
 DEFINITION to DnaJ protein.  
 ACCESSION AK095386  
 VERSION AK095386.1 GI:21754632  
 KEYWORDS oligo capping, f1s (full insert sequence).  
 SOURCE Homo sapiens (human)

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1  
 AUTHORS Tanigami,A., Fujiwara,T., Shibahara,T., Goto,Y., Hirao,M., Shimizu,F., Wakebe,H., Ono,T., Hishigaki,H., Watanabe,T., Ozaki,K., Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T., Kikuchi,H., Kanda,K., Matsuda,H., Magatsuna,M., Nakamura,Y., Sekine,M., Takahashi-Fujii,A., Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S., Nagahara,K., Maehuo,Y., Nagai,K. and Isogai,T.  
 TITLE NEDO human cDNA sequencing project  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 3012)  
 AUTHORS Isogai,T. and Yamamoto,J.  
 TITLE Direct Submission  
 JOURNAL Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team) ; 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan  
 COMMENT (E-mail:genominfo@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986) NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB) ; cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.  
 FEATURES  
 source  
 1..3012  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CTONG2015345"  
 /issue\_type="tongue, tumor tissue"  
 /clone\_lib="CTONG2"  
 /note="Cloning vector: pME18SFL3"  
 ORIGIN  
 Query Match 57.4%; Score 1845.2; DB 9; Length 3012;  
 Best Local Similarity 99.8%; Pred. No. 0;  
 Matches 1858; Conservative 0; Mismatches 3; Indels 1; Gaps 1;  
 QY 1338 GTCCGCGTATCAGAGATCTGTCCGAAGATTAATGAGCAGAAAGTCCAGAAATGGGC 1397  
 Db 1152 GTCCGCGTATCAGAGATCTGTCCGAAGATTAATGAGCAGAAAGTCCAGAAATGGGC 1211  
 QY 1398 CTCATCATGTCATGCTGTAACGGGAAGTTGTCAATGACAAAGCAGAAAGCGAG 1457  
 Db 1212 CTCATCATGTCATGCTGTAACGGGAAGTTGTCAATGACAAAGCAGAAAGCGAG 1271  
 QY 1458 AAGGTGAAGTATGACGTGACCTGCGCCGCGAGTGGCTGTGAAGACCTCGAAGCTC 1517  
 Db 1272 AAGGTGAAGTATGACGTGACCTGCGCCGCGAGTGGCTGTGAAGACCTCGAAGCTC 1331  
 QY 1518 ATTCCTCAGGAGGCTCCAGAGCTGGGCTGCTGCTTTATGACCCGCTGTGGGGGAA 1577  
 Db 1332 ATTCCTCAGGAGGCTCCAGAGCTGGGCTGCTGCTTTATGACCCGCTGTGGGGGAA 1391  
 QY 1578 GAGAAAGCTGAAGTGTCTATACATGTCGGGGGCTCTGATCAGGTGATGTCGTG 1637  
 Db 1392 GAGAAAGCTGAAGTGTCTATACATGTCGGGGGCTCTGATCAGGTGATGTCGTG 1451  
 QY 1638 GACAGTGAAGCCCTCCGATATCAAAAGAGTCCCAAGATGATGATGATGATTAAC 1697  
 Db 1452 GACAGTGAAGCCCTCCGATATCAAAAGAGTCCCAAGATGATGATGATGATTAAC 1511  
 QY 1698 GCCAAGAACCAATTTTAAAGGCGCAAAATCTTTCTGGAGGCTCAAAATTTG 1757  
 Db 1512 GCCAAGAACCAATTTTAAAGGCGCAAAATCTTTCTGGAGGCTCAAAATTTG 1571  
 QY 1758 GAAATGAAAAAACCAGACATGATGTTTATTTATTTATTTATTTATGAAAGGTGT 1817





[http://www.sanger.ac.uk/Projects/C\\_elegans/wormped](http://www.sanger.ac.uk/Projects/C_elegans/wormped) This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr1> RPI-126A5 is from the library RPCI-1 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm> VECTOR: pCVPAC2

This sequence is the entire insert of clone RPI-126A5.

## FEATURES

## source

Location/Qualifiers

1. .112158

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="RZPD:RZCP704A05126"

/db\_xref="taxon:9606"

/chromosome="1"

/map="p36.21-36.33"

/clone="RPI-126A5"

/clone\_1lb="RPCI-1"

51. .473

/note="match: GSS: Em:AQ655821"

64. .473

/note="match: GSS: Em:AQ168580"

178. .413

/note="MIR repeat: matches 1. .237 of consensus"

474. .954

/note="LRR7 repeat: matches 1. .450 of consensus"

1126. .1241

/note="L1PAB repeat: matches 6050. .6163 of consensus"

1572. .1861

/note="AluJo repeat: matches 1. .299 of consensus"

1896. .1953

/note="29 copies 2 mer ta 864 conserved"

1962. .2225

/note="AluSg repeat: matches 1. .261 of consensus"

2797. .3075

/note="AluJ repeat: matches 1. .294 of consensus"

3194. .3560

/note="MER1B repeat: matches 1. .333 of consensus"

4834. .4932

/note="MIR repeat: matches 56. .161 of consensus"

4966. .5046

/note="MIR repeat: matches 28. .109 of consensus"

5047. .5375

/note="AluJb repeat: matches 1. .310 of consensus"

5376. .5534

/note="MIR repeat: matches 109. .262 of consensus"

5595. .5722

/note="AluJo/FLAM repeat: matches 11. .132 of consensus"

5723. .5888

/note="AluJ repeat: matches 140. .300 of consensus"

5889. .6201

/note="AluJ repeat: matches 1. .303 of consensus"

6202. .6221

/note="AluJ repeat: matches 121. .140 of consensus"

6274. .6546

/note="LRR36 repeat: matches 294. .612 of consensus"

6581. .6748

/note="LRR36 repeat: matches 16. .204 of consensus"

6749. .7050

/note="AluSx repeat: matches 1. .305 of consensus"

7051. .7066

/note="LRR36 repeat: matches 3. .16 of consensus"

7067. .7203

/note="AluJo/FRAM repeat: matches 188. .310 of consensus"

7204. .7512

/note="AluYb8 repeat: matches 1. .314 of consensus"

7672. .7974

/note="AluSx repeat: matches 1. .297 of consensus"

7975. .8275

/note="AluSp repeat: matches 1. .301 of consensus"

8786. .9249

misc\_feature

/note="match: GSS: Em:AQ667975"  
8884. .9386  
/note="match: GSS: Em:AQ633106"  
9563. .9919  
/note="LIM2 repeat: matches 1151. .1479 of consensus"  
10039. .10240  
/note="AluJ repeat: matches 3. .204 of consensus"  
10814. .10915  
/note="MIR repeat: matches 88. .196 of consensus"  
11096. .11369  
/note="AluSp repeat: matches 11. .298 of consensus"  
11436. .11849  
/note="L2 repeat: matches 1816. .2259 of consensus"  
11820. .12011  
/note="L2 repeat: matches 2548. .2746 of consensus"  
13229. .13530  
/note="AluJ repeat: matches 1. .301 of consensus"  
14149. .14510  
/note="MT1I repeat: matches 3. .394 of consensus"  
14511. .14563  
/note="Char1el repeat: matches 2707. .2761 of consensus"  
14666. .15064  
/note="Char1el repeat: matches 169. .2729 of consensus"  
15183. .15340  
/note="match: GSS: Em:B00080"  
15480. .15650  
/note="Char1el repeat: matches 1. .175 of consensus"  
15716. .15899  
/note="AluJo repeat: matches 1. .192 of consensus"  
15900. .16009  
/note="LIM4 repeat: matches 2718. .2828 of consensus"  
16504. .16806  
/note="AluSx repeat: matches 1. .302 of consensus"  
17059. .17138  
/note="LIMB3A repeat: matches 6080. .6159 of consensus"  
17253. .17416  
/note="match: GSS: Em:AQ192257"  
17253. .17416  
/note="MIR repeat: matches 90. .260 of consensus"  
17405. .17625  
/note="MIR repeat: matches 15. .251 of consensus"  
17970. .18239  
/note="AluSg repeat: matches 1. .271 of consensus"  
17970. .18239  
/note="match: GSS: Em:AQ811010"  
19848. .66717  
/gene="d1126A5.1"  
/join<19848. .19977,22418. .22491,33075. .33176,54907. .55096,55759. .55844,56211. .56327,60828. .60983,62521. .62590,63381. .63438,63545. .63687,64639. .64768,65185. .66717)/gene="d1126A5.1"  
/product="d1126A5.1.1 (novel DnaJ domain protein)"  
/note="variant 1  
mostly supported by EGENES  
match: CDNA8: Em:AF052158 Em:AF086370  
match: ESTs: Em:AA386499 Em:D30988 Em:W57051 Em:A1560442 Em:AA529628 Em:A1007045 Em:R28618 Em:W7439 Em:AA85217 Em:AA31688 Em:AA299970 Em:AA537012 Em:AA589657 Em:W58891 Em:AU044630 Em:R57617 Em:AA162962 Em:AA359216 Em:AA681905 Em:AA046839 Em:AA047010 Em:R60577 Em:W87554 Em:A1017434 Em:AA039966 Em:A1123786 Em:AA039881 Em:AA069571 Em:W76508 Em:W71994 Em:H26530 Em:AA649997 Em:A1382155 Em:A1057398 Em:T79990 Em:R25837 Em:H10997 Em:R60630 Em:RS3484 Em:AA029772 Em:T12221 Em:Z45024 Em:A1280130 Em:A1521797 Em:A1146735 Em:H06388 Em:R16372 Em:H06281 Em:AA310656 Em:R72562 Em:AA057673 Em:A1078071 Em:A1245359 Em:AA029771 Em:A1556478 Em:AA379004 Em:AA967329 Em:AU016908 Em:A1475001 Em:AA166041 Em:AA207909 Em:AA144155 Em:AA616083 Em:AA563047"  
/evidence="not\_experimental  
20520. .20530  
/note="LIMB4 repeat: matches 6008. .6017 of consensus"  
complement(21198. .21671)







Dd	102065	CCGAGCGGGGAGGAGGTGTGTGTGATGATGAGCTGGGCCCCCCGGGCTTCCCGTGCGCTCT	102122
Qy	3109	GCGCAGAGTGCTTTCTGGGCACCCCATCTTCGTTCATTGTGCAGTGCACTGTACGAAG	3168
Dd	102123	GCGCAGAGGTCTTCTGGGCACCCTATCTCTGCGTTTCATTGTGCAGTGCACTGTACGAAG	102182
Qy	3169	GCACTCACCAATAAATACCTTCTCTGAAGCAGA	3202
Dd	102183	GCACTCACCAATAAATACCTTCTCTGAAGCAGA	102216
RESULT 13			
LOCUS	AX834019/c		
DEFINITION	Sequence 1143 from Patent EP1347046.		
ACCESSION	AX834019		
VERSION	AX834019.1	GI:39920154	
KEYWORDS			
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 Isogai,T., Sugiyama,T., Otsuka,T., Makamatsu,A., Sato,H., Ishii,S., Yamamoto,J.I., Isono,Y., Hiro,Y., Otsuka,K., Nagai,K., Irie,R., Tamechika,I., Seki,N., Yoshikawa,T., Otsuka,M., Nagahari,K. and Masuno,Y.		
TITLE	Full-length cDNA sequences		
JOURNAL	Patent: EP 1347046-A 1143 24-SEP-2003;		
FEATURES	Research Association for Biotechnology (JP)		
source	1..2052 /organism="Homo sapiens" /mol_type="unassigned DNA" /db_xref="taxon:9606"		
ORIGIN			
Query_Match	43.5%;	Score 1397.4;	DB 6; Length 2052;
Best Local Similarity	99.5%;	Pred. No. 0;	
Matches 1112;	Conservative 0;	Mismatches 6;	Indels 1; Gaps 1;
Qy	1784	TTTTTATTTTATTTATTTATTTATTAAGAAAGTGTGATCAATTATCAATTATGTGAAGGACA	1843
Dd	2052	GTTTTTATTTTATTTATTTATTTATTTATTAAGAAAGTGTGATCAATTATCAATTATGTGAAGGACA	1993
Qy	1844	TGCAGAACCCCAAGCTTTTGAAGGAGTCTGGGGGTAGACTGAGAGCGCCCACTGGGAAC	1903
Dd	1992	TGCAGAACCCCAAGCTTTTGAAGGAGTCTGGGGGTAGACTGAGAGCGCCCACTGGGAAC	1933
Qy	1904	CAGACTGACGCTGGGCCCATGGCTGTTTTTCCCAAGATCAAGTTCCTGAGAGGAAAGGGCTTC	1963
Dd	1932	CAGACTGACGCTGGGCCCATGGCTGTTTTTCCCAAGATCAAGTTCCTGAGAGGAAAGGGCTTC	1873
Qy	1964	TGGCCCTGACTCCGCTGTGTCCCGAGACACGTGTGACACCGACCCGCGCCCTTAGT	2023
Dd	1872	TGGCCCTGACTCCGCTGTGTCCCGAGACACGTGTGACACCGACCCGCGCCCTTAGT	1813
Qy	2024	TCTTGGCTGGGCTGTGAGAGGTGTCTGTGAGCACCTTGCCCTTCAACAAGAGCGTAGGCC	2083
Dd	1812	TCTTGGCTGGGCTGTGAGAGGTGTCTGTGAGCACCTTGCCCTTCAACAAGAGCGTAGGCC	1753
Qy	2084	ACTTGTGACGTACAGGCTGAACATGGGAAACAACCTGAAAAGCAGGCAAGCGCTCCCGGTC	2143
Dd	1752	ACTTGTGACGTACAGGCTGAACATGGGAAACAACCTGAAAAGCAGGCAAGCGCTCCCGGTC	1693
Qy	2144	AGGAGCGCTTGCTGTGCTGTGCTTCCCATGACCACTCTCTCTGCTGCAATATTACTGCT	2203
Dd	1692	AGGAGCGCTTGCTGTGCTGTGCTTCCCATGACCACTCTCTCTGCTGCAATATTACTGCT	1633
Qy	2204	TGAATCTGAGAGCAATATGGGGGTTTATTAACAATCTGCTTTTATCTGGAACAACGGGTTT	2263
Dd	1632	TGAATCTGAGAGCAATATGGGGGTTTATTAACAATCTGCTTTTATCTGGAACAACGGGTTT	1573

OY		2264	GGAATTATAGTGTCTTTTTCACCACCTCCAGACTGTCAAAGCATTTCCACCGGGCCCC	2263
Db		1572	GGAAATTATAGTGTCTTTTTCACCACCTCCAGACTGTCAAATCAATTCACCGGGCCCC	1513
OY		2234	TGGCTTGGAAGAAGGATGATGTAACTCCGCATCCAGGGGCTTAGGCCCTGACACAGATGGC	2289
Db		1512	TGGCTTGGAAGAAGGATGATGTAACTCCGCATCCAGGGGCTTAGGCCCTGACACAGATGGC	1455
OY		2384	TTCCCGTATCCCGGTGGAAAAACGCTTCGACCAAGCGGGCTTAGAGCTGTGTCTCC	2443
Db		1452	TTCCCGTATCCCGGTGGAAAAACGCTTCGACCAAGGGGGCTTAGAGCTGTGTGTCTCC	1399
OY		2444	TCACACCCCTGACACAACCACTTCAAAGTGTCACTGCTGGGCAAGGGCAGCTTAAGAAGA	2503
Db		1392	TCACCTGCTCACACCACTCCTCAGAGTGTCAAGTGTGGCAAGGGCAGCTCAAGAAGA	1333
OY		2504	CAGAACCAAGGGCTTGGCAAGCATTCAGACACACCAACCCAAAGGGGTGACCCCAAGGC	2563
Db		1332	CAGAACCAAGGGCTTGGCAAGCATTCAGACACACCAACCCAAAGGGGTGACCCCAAGGC	1279
OY		2564	CCGGCCCTGTGTACCCACAGCAGTGTGGCACTGAGCTCCCGCTCTCTGACAGTCCACGCTCC	2623
Db		1272	CCGGCCCTGTGTACCCACAGCAGTGTGGCACTGAGCTCCCGCTCTCTGACAGTCCACGCTCC	1213
OY		2624	TCACAGAACACCAAGGCTGTGTCTCCGGAACCTTCTTCAGACCTTCTTCACAGTCCC	2683
Db		1212	TCACAGAACACCAAGGCTGTGTCTCCGGAACCTTCTTCAGACCTTCTTCACAGTCCC	1153
OY		2684	CACTTGGGATACAAATGTCAGCGGAGCTAAGAACCCCTCCACCGGCTGTGACACTGTGGCTGC	2743
Db		1152	CACTTGGGATACAAATGTCAGCGGAGCTAAGAACCCCTCCACCGGCTGTGACACTGTGGCTGC	1093
OY		2744	AGTAAAGTTAAGTAGAGGCTGTCTCTCGGGGCTGTGAAGTGTGCAGCATCAGTTGCTCTT	2803
Db		1092	AGTAAAGTTAAGTAGAGGCTGTCTCTCGGGGCTGTGAAGTGTGCAGCATCAGTTGCTCTT	1033
OY		2804	GCTGACCCCTCGAAGCAAGCGCGCACAGAGTGTGTGTGAGACAGCTGGCGGGGGGGCC	2863
Db		1032	GCTGACCCCTCGAAGCAAGCGCGCACAGAGTGTGTGTGAGACAGCTGGCGGGGGGGCC	973
OY		2864	CCAAGCTGGCGGGGCTCCAGGCCACCCACAGCTGTGTGTGAATTCAGGCCAATCTCCC	2923
Db		972	CCAAGCTGGCGGGGCTCCAGGCCACCCACAGCTGTGTGTGAATTCAGGCCAATCTCCC	913
OY		2924	AGCACTGGTATCTGATGTAACGGCTAAGAACCTCTCTCTCTGTTTTTGAAGAAAGAGTTGC	2983
Db		912	AGCACTGGTATCTGATGTAACGGCTAAGAACCTCTCTCTCTGTTTTTGAAGAAAGAGTTGC	853
OY		2984	GATTGTCCAATTCTGTAACTTCATCTCCATTTTAAAAGGTTTTCTGACGGGCC	3043
Db		852	GATTGTCCAATTCTGTAACTTCATCTCCATTTTAAAAGGTTTTCTGACGGGCC	794
OY		3044	ACGGCCCGAAGCGCGGTGAAGGCTGTGTGTGTGATAGCTTGGGCCCCCGGGCTTCCCGTGC	3103
Db		793	ACGGCCCGAAGCGCGGTGAAGGCTGTGTGTGTGATAGCTTGGGCCCCCGGGCTTCCCGTGC	734
OY		3104	CCTGTGCGCAGAGGCTTCTGGGGCAACCATCTGTGAGTTTCAATTTGACGTGACCTGTAC	3163
Db		733	CCTGTGCGCAGAGGCTTCTGGGGCAACCATCTGTGAGTTTCAATTTGACGTGACCTGTAC	674
OY		3164	AGAAAGCACTCACCAATATAACCTTCTCTGAAGACAGA	3202
Db		673	AGAAAGCACTCACCAATATAACCTTCTCTGAAGACAGA	635
<hr/>				
RESULT 14				
AKO96217/c				
LOCUS	AKO96217	2052 bp	mRNA	linear
DEFINITION	Homo sapiens cDNA FLJ38898 f1s, clone NT2ME2000374.	PRI	15-JUL-2002	
ACCESSION	AKO96217			
VERSION	AKO96217.1	GI:21755650		
KEYWORDS	oligo capping; f1s (full insert sequence).			
SOURCE	Homo sapiens (human)			

ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
AUTHORS 1  
Ninomiya, K., Magatsuma, M., Kanda, K., Kondo, H., Yokoi, T., Kodaira, H., Furuya, T., Takahashi, M., Kikkawa, E., Omura, Y., Abe, K., Kamihara, K., Katsuta, N., Sato, K., Tanikawa, M., Yamazaki, M., Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T., Yamura, K., Yamashita, H., Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Murakawa, K., Kanehori, K., Takahashi, Fujii, A., Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Nagahara, K., Masuno, Y., Nagai, K. and Isogai, T.  
TITLE Unpublished  
JOURNAL NEDO human cDNA sequencing project  
REFERENCE 2 (bases 1 to 2052)  
AUTHORS Isogai, T. and Yamamoto, J.  
TITLE Direct Submission  
JOURNAL Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team) ; 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan  
COMMENT (E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986) NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan, cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.  
FEATURES  
Source  
1. .2052  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="NT2NE200374"  
/cell\_line="NT2"  
/cell\_type="teratocarcinoma"  
/clone\_lib="NT2NE2"  
/note="Cloning vector: pME18SFL3-mRNA from NT2 neuron after the differentiation of NT2 neuronal precursor cells."  
28. .555  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="BAC04727.1"  
/db\_xref="GI:21755651"  
/translation="MPKSCAAROCNRYSSRRKQLTFHRPPSPRELLKEMVINTGRG NPKPKHTVISCSEHPRPBCSAGRNKLNKHAVPVPAFODPTQVRENTDPAISRG MASSSQREKTSFKRSQVLPBAGAGBDSFGKMDTLALEIQLPENAGHVAKQMLFNVE NGTPASREALMISEE"  
ORIGIN  
Query Match 43.5%; Score 1397.4; DB 9; Length 2052;  
Best Local Similarity 99.5%; Pred. No. 0;  
Matches 1412; Conservative 0; Mismatches 6; Indels 1; Gaps 1;  
QY 1784 GTTTATTTTATTTATTTATTTATAGAGGTGTACATTATCAATTATGTGAAGGACA 1843  
DB 2052 GTTTTATTTTATTTATTTATTTATAGAGGTGTACATTATCAATTATGTGAAGGACA 1993  
QY 1844 TGACAGACCCCGAGCTTTGAGGGTGTGGGGGTAGAGCTGAGGACGCCCACTGGGAAC 1903  
DB 1992 TGACAGACCCCGAGCTTTGAGGGTGTGGGGGTAGAGCTGAGGACGCCCACTGGGAAC 1933  
QY 1904 CAGACTGACGCGGCGCATGTGCTTTCCCAAGATCAGTCCCTGAGAGGGAAGGGCTC 1963  
DB 1932 CAGACTGACGCGGCGCATGTGCTTTCCCAAGATCAGTCCCTGAGAGGGAAGGGCTC 1873  
QY 1964 TGGCCCTGACTCCGCTGTGTCTCCGAGACACAGTGTGACCGGACGCCGCGCTGTAGT 2023  
DB 1872 TGGCCCTGACTCCGCTGTGTCTCCGAGACACAGTGTGACCGGACGCCGCGCTGTAGT 1813  
QY 2024 TCTTGCGTGGGTCTGAGAGGTGTCTGTGAGACACCTGCGCTCAACAAGAGCGTGAAC 2083

DB 1812 TCTTGCGTGGGTCTGAGAGGTGTCTGTGAGACACCTGCGCTCAACAAGAGCGTGAAC 1753  
QY 2084 ACTTCGACGTCACGCTGACATGAGGAACCAACCTGAAGAGAGAGGCTCCCGGTC 2143  
DB 1752 ACTTCGACGTCACGCTGACATGAGGAACCAACCTGAAGAGAGAGGCTCCCGGTC 1693  
QY 2144 AGGAGACCTCTGCTGTGTGTGCTGCTCCATGACCACTCTCTGTGTAATTAATCTGCT 2203  
DB 1692 AGGAGACCTCTGCTGTGTGTGCTGCTCCATGACCACTCTCTGTGTAATTAATCTGCT 1633  
QY 2204 TGAATGTGAGACAGATTCGGGTTTAAATCTGCTTTTATCTGAAACAAACGGGTTT 2263  
DB 1632 TGAATGTGAGACAGATTCGGGTTTAAATCTGCTTTTATCTGAAACAAACGGGTTT 1573  
QY 2264 GGAATTAATGCTGCTTTTCCCACTCCAGAGGCTCAAGTCAATTCACCGGCCCC 2323  
DB 1572 GGAATTAATGCTGCTTTTCCCACTCCAGAGGCTCAATTAATTCACCGGCCCC 1513  
QY 2324 TGGCTTGGAGACAGGTAAGTGAATCCGATCCAGAGGCTTAGGCTGACACAGGTGAC 2383  
DB 1512 TGGCTTGGAGACAGGTAAGTGAATCCGATCCAGAGGCTTAGGCTGACACAGGTGAC 1453  
QY 2384 TTCCGCTATCCCGTGGGAAAAAGCCCTGACACAGCGGCTTAGGCTGTGTCTCC 2443  
DB 1452 TTCCGCTATCCCGTGGGAAAAAGCCCTGACACAGCGGCTTAGGCTGTGTCTCC 1393  
QY 2444 TCCACGCGCTGACACACCCACCTCCAGAGTGAAGTGTGGGACGAGGCTCAAGAGA 2503  
DB 1392 TCCACGCGCTGACACACCCACCTCCAGAGTGAAGTGTGGGACGAGGCTCAAGAGA 1333  
QY 2504 CAGACACAGACGCTTGGCAAGACATCAGACACCAACCCAAAGCGGTGACCCCAAGC 2563  
DB 1332 CAGACACAGACGCTTGGCAAGACATCAGACACCAACCCAAAGCGGTGACCCCAAGC 1273  
QY 2564 CCGGCGCGTGTATCCAGAGAGTGTGACAGTCTCCGCTCTGACAGTCCAGGCTCC 2623  
DB 1272 CCGGCGCGTGTATCCAGAGAGTGTGACAGTCTCCGCTCTGACAGTCCAGGCTCC 1213  
QY 2624 TCACAGGAACACAGGAGCTGTGCTCCGAGGCTTCTTCAGACCTTCTCCACAGTGC 2683  
DB 1212 TCACAGGAACACAGGAGCTGTGCTCCGAGGCTTCTTCAGACCTTCTCCACAGTGC 1153  
QY 2684 CACTTGGATGACAGATGACGAGGACTAGACCCCTTCACAGGCTGACCTCGCTGC 2743  
DB 1152 CACTTGGATGACAGATGACGAGGACTAGACCCCTTCACAGGCTGACCTCGCTGC 1093  
QY 2744 AGTAAAGTTAGCTGAGGCTGTCTCTCGGAGGCTTGAAGTGGCAGCATCACTTCTCT 2803  
DB 1092 AGTAAAGTTAGCTGAGGCTGTCTCTCGGAGGCTTGAAGTGGCAGCATCACTTCTCT 1033  
QY 2804 GGTGACCCCTGGAGACAGCGCGGACAGGAGTGTGAGACAGTGGCGGCGGAGGCTC 2863  
DB 1032 GGTGACCCCTGGAGACAGCGCGGACAGGAGTGTGAGACAGTGGCGGCGGAGGCTC 973  
QY 2864 CCAAGCTGCGCGGCTCCAGGCCCAACCCACAGCTGTGTGAGAGTCAAGCCAACTCCCC 2923  
DB 972 CCAAGCTGCGCGGCTCCAGGCCCAACCCACAGCTGTGTGAGAGTCAAGCCAACTCCCC 913  
QY 2924 AGCAGTGTATCTGAGTAAGGCTTAAGAACTCTCTCTGTGTTTGAAGAGAGTGTG 2983  
DB 912 AGCAGTGTATCTGAGTAAGGCTTAAGAACTCTCTCTGTGTTTGAAGAGAGTGTG 853  
QY 2984 GGTGTCACAACTCTGTAACATTCATCTCAATTTTAAAAAGTTTCTCTGACGGCCCC 3043  
DB 852 GGTGTCACAACTCTGTAACATTCATCTCAATTTT -AAAAAGTTTCTCTGACGGCCCC 794  
QY 3044 ACGGCCGAGCGCGGTGAGCGTGTGTGACAGAGCTGTGGCCCGCGGCTTCCGCTGCG 3103  
DB 793 ACGGCCGAGCGCGGTGAGCGTGTGTGACAGAGCTGTGGCCCGCGGCTTCCGCTGCG 734  
QY 3104 CCTTGCAGAGGTCTTGTGGGACCACTCTCTGCGTTTCAATTGACAGTGAAGTGAAC 3163

Db 733 CCTCTGCGAGGTGTTCTTGCGACCACTCTGCGTTTCATTGTGAGTGCAGCTGAC 674  
Qy 3164 AGAAGGACCTCACCACATTAACCTTCTCTGAAGCAGA 3202  
Db 673 AGAAGGACCTCACCACATTAACCTTCTCTGAAGCAGA 635

RESULT 15  
AC062024/c 185061 bp DNA linear HTG 01-SEP-2000  
LOCUS Homo sapiens chromosome 1 clone RP11-239H20, WORKING DRAFT  
DEFINITION AC062024  
ACCESSION AC062024 2 GI:9958200  
VERSION HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
KEYWORDS Homo sapiens (human)  
SOURCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
ORGANISM

REFERENCE 1 (bases 1 to 185061)  
AUTHORS Waterston,R.H.  
TITLE The sequence of Homo sapiens clone  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 185061)  
AUTHORS Waterston,R.H.  
TITLE Direct Submission  
JOURNAL Submitted (21-APR-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
COMMENT ON Sep 1, 2000 this sequence version replaced gi:7630906.

Center: Washington University Genome Sequencing Center  
Web site: http://genome.wustl.edu/gsc/index.shtml  
Project Information  
Center project name: H.NH0239H20  
Summary Statistics  
Sequencing vector: M13; 100%  
Chemistry: Dye-terminator Big Dye; 0% of reads  
Chemistry: Dye-terminator Big Dye; 0% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 16851 bases at least Q40  
Consensus quality: 174217 bases at least Q30  
Consensus quality: 177307 bases at least Q20  
Insert size: 173000; agarose-fp  
Insert size: 182161; sum-of-contigs  
Quality coverage: 3.82 in Q20 bases; agarose-fp  
Quality coverage: 3.68 in Q20 bases; sum-of-contigs  
NOTE: This is a 'working draft' sequence. It currently consists of 30 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of 'N', but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 1990: contig of 1990 bp in length  
1 1991 2090: gap of unknown length  
\* 2091 3876: contig of 1786 bp in length  
\* 3877 3976: gap of unknown length  
\* 3977 5964: contig of 1988 bp in length  
\* 5965 6064: gap of unknown length  
\* 6065 7612: contig of 1548 bp in length  
\* 7613 7712: gap of unknown length  
\* 7713 11594: contig of 3882 bp in length  
\* 11595 11694: gap of unknown length  
\* 11695 15489: contig of 3795 bp in length  
\* 15490 15589: gap of unknown length  
\* 15590 19085: contig of 3496 bp in length  
\* 19086 19185: gap of unknown length

19186 21398: contig of 2213 bp in length  
\* 21399 21498: gap of unknown length  
\* 21499 33593: contig of 12095 bp in length  
\* 33594 33693: gap of unknown length  
\* 33694 36499: contig of 2806 bp in length  
\* 36500 36599: gap of unknown length  
\* 36600 40210: contig of 3611 bp in length  
\* 40211 40311: gap of unknown length  
\* 40311 43308: contig of 2998 bp in length  
\* 43309 43408: gap of unknown length  
\* 43409 46757: contig of 3349 bp in length  
\* 46758 46857: gap of unknown length  
\* 46858 51135: contig of 4278 bp in length  
\* 51136 51235: gap of unknown length  
\* 51236 55293: contig of 4058 bp in length  
\* 55294 55393: gap of unknown length  
\* 55394 62010: contig of 6617 bp in length  
\* 62011 62111: gap of unknown length  
\* 62111 68999: contig of 6889 bp in length  
\* 69000 69099: gap of unknown length  
\* 69100 74515: contig of 5416 bp in length  
\* 74516 74615: gap of unknown length  
\* 74616 80987: contig of 6372 bp in length  
\* 80988 81087: gap of unknown length  
\* 81088 87983: contig of 6896 bp in length  
\* 87984 88083: gap of unknown length  
\* 88084 98428: contig of 10345 bp in length  
\* 98429 98528: gap of unknown length  
\* 98529 107911: contig of 9383 bp in length  
\* 107912 108011: gap of unknown length  
\* 108012 117076: contig of 9065 bp in length  
\* 117077 117176: gap of unknown length  
\* 117177 128937: contig of 11761 bp in length  
\* 128938 129037: gap of unknown length  
\* 129038 138066: contig of 9029 bp in length  
\* 138067 138166: gap of unknown length  
\* 138167 149334: contig of 11168 bp in length  
\* 149335 149434: gap of unknown length  
\* 149435 165184: contig of 15750 bp in length  
\* 165185 165284: gap of unknown length  
\* 165285 181704: contig of 16420 bp in length  
\* 181705 181804: gap of unknown length  
\* 181805 183334: contig of 1530 bp in length  
\* 183335 183434: gap of unknown length  
\* 183435 185061: contig of 1627 bp in length.  
Location/Qualifiers  
1. 185061  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/clone="RP11-239H20"  
1. 1990  
/note="assembly\_name:Contig11"  
2091. 3876  
/note="assembly\_name:Contig12"  
3977. 5964  
/note="assembly\_name:Contig13"  
6065. 7612  
/note="assembly\_name:Contig14"  
7713. 11594  
/note="assembly\_name:Contig15"  
11695. 15489  
/note="assembly\_name:Contig16"  
15590. 19085  
/note="assembly\_name:Contig17"  
19186. 21398  
/note="assembly\_name:Contig18  
Clone\_end:SP6  
vector\_side:left"  
21499. 33593  
/note="assembly\_name:Contig36"  
33694. 36499

/note="assembly\_name:Contig19"  
36600..40210  
/note="assembly\_name:Contig20"  
40311..43308  
/note="assembly\_name:Contig21"  
43409..46757  
/note="assembly\_name:Contig22"  
46858..51135  
/note="assembly\_name:Contig23"  
51236..55293  
/note="assembly\_name:Contig24"  
55394..62010  
/note="assembly\_name:Contig25"  
62111..68999  
/note="assembly\_name:Contig26"  
69100..74515  
/note="assembly\_name:Contig27"  
74616..80987  
/note="assembly\_name:Contig28"  
81088..87983  
/note="assembly\_name:Contig29"  
88084..98428  
/note="assembly\_name:Contig30"  
98529..107911  
/note="assembly\_name:Contig31"  
108012..117076  
/note="assembly\_name:Contig32"  
117177..128937  
/note="assembly\_name:Contig33"  
129038..138066  
/note="assembly\_name:Contig34"  
138167..149334  
/note="assembly\_name:Contig35"  
149435..165184  
/note="assembly\_name:Contig37"  
165285..181704  
/note="assembly\_name:Contig38"  
181805..183334  
/note="assembly\_name:Contig8"  
183435..185061  
/note="assembly\_name:Contig9"

ORIGIN

Query Match 38.5%; Score 1236.4; DB 2; Length 185061;  
Best Local Similarity 92.1%; Pred. No. 0;  
Matches 1273; Conservative 0; Mismatches 106; Indels 3; Gaps 3;  
QY 1821 ATTATCAATTATGTGAAGGACATGCAGACACCCAGCTTTTGAAGGTGCTGGGGGTAG 1880  
DB 51236 ANNN 51177  
QY 1881 ACTGAGCAGCCCACTGGGAACAAGACTGCAGCTGGCCCATGGCTTTTCCCAAGGA 1940  
DB 51176 NNN 51117  
QY 1941 TCAGTTCCTGAGGAGGAGGCTCTGGCCCTGACTCCGCTGTCTCCGAGACACAGTGTG 2000  
DB 51116 TCAGTTCCTGAGGAGGAGGCTCTGGCCCTGACTCCGCTGTCTCCGAGACACAGTGTG 51057  
QY 2001 ACCGCAAGCCCGCCCTGTAGTTCTTGGCTGGGTCTGAGAGTGTCTGTGAGACACCTTG 2060  
DB 51056 ACCGCAAGCCCGCCCTGTAGTTCTTGGCTGGGTCTGAGAGTGTCTGTGAGACACCTTG 50997  
QY 2061 CCCTCAACCAAGAGAGCTGAGCACTTCTGCACTCCACGCTGAAACATGGGAAACAACCTG 2120  
DB 50996 CCCTCAACCAAGAGAGCTGAGCACTTCTGCACTCCACGCTGAAACATGGGAAACAACCTG 50937  
QY 2121 AAAAGCAGGAGGCTCTCCGCTGAGGAGGCTCTGCTGTGTCTGGCTTCCATGACCACTT 2180  
DB 50936 AAAAGCAGGAGGCTCTCCGCTGCA-GGAGCTCTGCTGTGTCTGGCTTCCATGACCACTT 50878  
QY 2181 CCTCTGCTGAATATTACTGCTTGAATCTGAGACAGATTGCGGGTTTATAAAACTGCTT 2240

DB 50877 CCTCTGCTGAATATTACTGCTTGAATCTGAGACAGATTGCGGGTTTATAAAACTGCTT 50818  
QY 2241 TTTATCTGAGAACAAACGGGTTTGGAAATTAGTCGCTTTTCCCACTCCAGAGCTG 2300  
DB 50817 TTTATCTGAGAACAAACGGGTTTGGAAATTAGTCGCTTTTCCCACTCCAGAGCTG 50758  
QY 2301 CTCAGATCAATTCACCGGCCCCCTCGGCTTGGGACAGGGTAGTGTAACTCCGATCCAG 2360  
DB 50757 CTCAGATCAATTCACCGGCCCCCTCGGCTTGGGACAGGGTAGTGTAACTCCGATCCAG 50698  
QY 2361 GGCCTTAGCCTTACACAGGTGCTTCCGCTATCCCGGTGGGAAAACGCCCTTCCACACG 2420  
DB 50697 GGCCTTAGCCTTACACAGGTGCTTCCGCTATCCCGGTGGGAAAACGCCCTTCCACACG 50638  
QY 2421 GGGCTTAGCCTGAGCTGCTGCTCCCTCCACAGCCTGACCAACCCACCTCCAGAGTGC 2480  
DB 50637 GGGCTTAGCCTGAGCTGCTGCTCCCTCCACAGCCTGACCAACCCACCTCCAGAGTGC 50578  
QY 2481 TGGGCAAGGGCAGCTCAAGAGGACAGAACAGAGCGCTTGGCAAGACATCAGACACCCA 2540  
DB 50577 TGGGCAAGGGCAGCTCAAGAGGACAGAACAGAGCGCTTGGCAAGACATCAGACACCCA 50518  
QY 2541 ACCCAAGGCGTGAACCCCAAGGCCCGGCGCTGTGTATCCAGACAGGTGCACTGCAGCTC 2600  
DB 50517 ACCCAAGGCGTGAACCCCAAGGCCCGGCGCTGTGTATCCAGACAGGTGCACTGCAGCTC 50458  
QY 2601 CCGCTCTGACAGGTTCAGGCTGCTCAACAGAAACACAGGGCGCTGTGCTCCGAGCCTTCC 2660  
DB 50457 CCGCTCTGACAGGTTCAGGCTGCTCAACAGAAACACAGGGCGCTGTGCTCCGAGCCTTCC 50398  
QY 2661 TTTCAGACCTTCTCTCCACGCTGCCAATTGGAGTGAATGACAGCGGAGCTGAGACCCCC 2720  
DB 50397 TTTCAGACCTTCTCTCCACGCTGCCAATTGGAGTGAATGACAGCGGAGCTGAGACCCCC 50338  
QY 2721 TCACGCGGCTGAGACTCGGCTGCACTGAATTAAGTGAAGGCTGTCTCTCGGGGCTTGA 2780  
DB 50337 TCACGCGGCTGAGACTCGGCTGCACTGAATTAAGTGAAGGCTGTCTCTCGGGGCTTGA 50278  
QY 2781 AGTGGCAGCATCAAGTGTCTTGTGCTGACCCCTCGAGACAGCGCGCACAGGTGTGAC 2840  
DB 50277 AGTGGCAGCATCAAGTGTCTTGTGCTGACCCCTCGAGACAGCGCGCACAGGTGTGAC 50218  
QY 2841 TGAAGCAGCTGAGCGGGGGGGGCCCAAGCTGCGCGGCTCCAGCCCAACCACTGTT 2900  
DB 50217 TGAAGCAGCTGAGCGGGGGGGGCCCAAGCTGCGCGGCTCCAGCCCAACCACTGTT 50158  
QY 2901 GCTGAAGTGAAGCACAATCTCCCAAGCACTGTATCTGAGTAAACGGCTAAGAACCTTCCTC 2960  
DB 50157 GCTGAAGTGAAGCACAATCTCCCAAGCACTGTATCTGAGTAAACGGCTAAGAACCTTCCTC 50098  
QY 2961 CTCTGGTTTGAAGACAGTTCGGGTTGTCAATTCTGTAACATTCTCATTTTCTT 3020  
DB 50097 CTCTGGTTTGAAGACAGTTCGGGTTGTCAATTCTGTAACATTCTCATTTTCTT 50039  
QY 3021 AAAAAGTTTCTGACGAGCCCAAGGCCCAAGCGCGGTGAGCGTGTGATGAGC 3080  
DB 50038 AAAAAGTTTCTGACGAGCCCAAGGCCCAAGCGCGGTGAGCGTGTGATGAGC 49980  
QY 3081 CTGGGCCCCGGGCTTCCGCTGCTCTGCGCAGAGTGTCTTGGGACACCATCTCTGCG 3140  
DB 49979 CTGGGCCCCGGGCTTCCGCTGCTCTGCGCAGAGTGTCTTGGGACACCATCTCTGCG 49920  
QY 3141 GTTTCATTGCACTGACGTGACAGAGGCACTCAACAAATAAACCCTTCTGAAAGCA 3200  
DB 49919 GTTTCATTGCACTGACGTGACAGAGGCACTCAACAAATAAACCCTTCTGAAAGCA 49860  
QY 3201 GA 3202  
DB 49859 GA 49858



***This Page Blank (uspto)***





CC respectively. The sequences can be used to treat and prevent disorders  
CC associated with altered expression or activity of HGP comprising  
CC administering a composition comprising the polypeptide or an antagonist  
CC to a patient (claimed). The human chaperone proteins are also useful for  
CC the diagnosis, treatment or prevention of neurodegenerative, metabolic,  
CC developmental, autoimmune/inflammatory disorders and cell proliferative  
CC disorders including cancer

XX Sequence 3213 BP; 771 A; 876 C; 894 G; 672 T; 0 U; 0 Other;

Query Match 100.0%; Score 3213; DB 3; Length 3213;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 3213; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 GAAAGGTGGAGATGCGAGCGCTTGAACGAGAGAGCTGGACCAATGAAGACTATT 60
DB 1 GAAAGGTGGAGATGCGAGCGCTTGAACGAGAGAGCTGGACCAATGAAGACTATT 60
QY 61 ACTGTTGCTGAACTGTCGAGGAGGCTCTTCTGAAGAGCTGAAAGCTCTACCGGA 120
DB 61 ACTGTTGCTGAACTGTCGAGGAGGCTCTTCTGAAGAGCTGAAAGCTCTACCGGA 120
QY 121 GGTCTGTATGCTCTACCATCCAGACAGACCCAGAGCTCAAGTCAACAGCGG 180
DB 121 GGTCTGTATGCTCTACCATCCAGACAGACCCAGAGCTCAAGTCAACAGCGG 180
QY 181 AAGCACTGTTAACCCTGTTACCAAGGCTTATGAAGCTTAGTGAACCCCAACAGAG 240
DB 181 AAGCACTGTTAACCCTGTTACCAAGGCTTATGAAGCTTAGTGAACCCCAACAGAG 240
QY 241 CCATCTATGATATATATGAGAGAGAGCTGAAATGGAAGATGGAGTGTGAGAA 300
DB 241 CCATCTATGATATATATGAGAGAGAGCTGAAATGGAAGATGGAGTGTGAGAA 300
QY 301 GAGAGAGAACCCCTGCTGAAATTCGAGAGAGCTTTGAGCGCTGCAGAGAGAGAG 360
DB 301 GAGAGAGAACCCCTGCTGAAATTCGAGAGAGCTTTGAGCGCTGCAGAGAGAGAG 360
QY 361 AGAGAGATGTCAGAGCGAACCAATCCAAAGGAAAGATCAGCGTTGAGTATGCCA 420
DB 361 AGAGAGATGTCAGAGCGAACCAATCCAAAGGAAAGATCAGCGTTGAGTATGCCA 420
QY 421 CCGACCTTTTGTATCGCTATGATGAGAGATGAGATGTGTCCGAGATGATTCGCG 480
DB 421 CCGACCTTTTGTATCGCTATGATGAGAGATGAGATGTGTCCGAGATGATTCGCG 480
QY 481 AGATGAAATTAATAAATGCAATATCCAGTCCATTGAGGACCCCTTGAACGCGAC 540
DB 481 AGATGAAATTAATAAATGCAATATCCAGTCCATTGAGGACCCCTTGAACGCGAC 540
QY 541 ACAAGCATCTCTCTGGAAGGCTCTCAACCCAGATGGAATGAGAGGATTCATT 600
DB 541 ACAAGCATCTCTCTGGAAGGCTCTCAACCCAGATGGAATGAGAGGATTCATT 600
QY 601 ACTTTGGCTCAGACGAGTAATCTTGGCAAGGAGATGGAGAGTGGAAATTTGAGCT 660
DB 601 ACTTTGGCTCAGACGAGTAATCTTGGCAAGGAGATGGAGAGTGGAAATTTGAGCT 660
QY 661 GAGACCTACAGGGGCTTTGTGTGCTCAAGCTGTTCCTTAATCTCAACCAAGATGCT 720
DB 661 GAGACCTACAGGGGCTTTGTGTGCTCAAGCTGTTCCTTAATCTCAACCAAGATGCT 720
QY 721 TTGTGCAACAACTGTGCTCTGCAAGTTTCATCCGTGGAAATCCGACCCGCTACCA 780
DB 721 TTGTGCAACAACTGTGCTCTGCAAGTTTCATCCGTGGAAATCCGACCCGCTACCA 780
QY 781 CTGTCTAGCTCGAACTTAGACAAACCCGAGGCTACCTGACAGTGGCAATGGGAT 840
DB 781 CTGTCTAGCTCGAACTTAGACAAACCCGAGGCTACCTGACAGTGGCAATGGGAT 840
QY 841 TCCAGTACGCAATGAACTAGCATGTGTGAGACACTTAAACCAAGCACTTCACTGTG 900
DB 841 TCCAGTACGCAATGAACTAGCATGTGTGAGACACTTAAACCAAGCACTTCACTGTG 900
```

```
QY 901 CCCTGAGCTGGGAATCCCTCACTCCTTTGCACTGATCAGTATGACACAAATTCAG 960
DB 901 CCCTGAGCTGGGAATCCCTCACTCCTTTGCACTGATCAGTATGACACAAATTCAG 960
QY 961 ATGACGATCAGACTGTGTGAAAGATCCCTCAAGCAGCTCTTTGGGACGTGTGG 1020
DB 961 ATGACGATCAGACTGTGTGAAAGATCCCTCAAGCAGCTCTTTGGGACGTGTGG 1020
QY 1021 AGTACGAGCTGAGAGAGATCTCCAGGACAGAGCTTTGGGTGCACTGTGAGCTTG 1080
DB 1021 AGTACGAGCTGAGAGAGATCTCCAGGACAGAGCTTTGGGTGCACTGTGAGCTTG 1080
QY 1081 GAGTTCCAGAGGCGCTTCTCAAAAGTCAAGCTCAACAGGCGCAGTCAAGTACTTCT 1140
DB 1081 GAGTTCCAGAGGCGCTTCTCAAAAGTCAAGCTCAACAGGCGCAGTCAAGTACTTCT 1140
QY 1141 TCCCTATTCATTGACGACCAAGCTTTCGACGCGCATGTTCTATGCAACCGTGGG 1200
DB 1141 TCCCTATTCATTGACGACCAAGCTTTCGACGCGCATGTTCTATGCAACCGTGGG 1200
QY 1201 CTCTAGTGTCTAATTTGCTATGACCGTCTGATCATCAACCATACCTCAGGCTCAGA 1260
DB 1201 CTCTAGTGTCTAATTTGCTATGACCGTCTGATCATCAACCATACCTCAGGCTCAGA 1260
QY 1261 AAGAGAAAGATTTGAGAGAGAGGAGAAAGCGCGCACCGATGTGTGAGAAAGAGC 1320
DB 1261 AAGAGAAAGATTTGAGAGAGAGGAGAAAGCGCGCACCGATGTGTGAGAAAGAGC 1320
QY 1321 AAGAGCGAGTCCGCTGTCCGCTGATGACAGAAATCTGTCCGAAGATTAATTGAGCAG 1380
DB 1321 AAGAGCGAGTCCGCTGTCCGCTGATGACAGAAATCTGTCCGAAGATTAATTGAGCAG 1380
QY 1381 AAGAGTCCGAATGGGCTCTCATCTGTCAATGCTTGTGAGGAGAGTTGTCAATGAGA 1440
DB 1381 AAGAGTCCGAATGGGCTCTCATCTGTCAATGCTTGTGAGGAGAGTTGTCAATGAGA 1440
QY 1441 AAGAGCAGAAAGCGAGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1500
DB 1441 AAGAGCAGAAAGCGAGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1500
QY 1501 TGAAGACTCGAAGCTCATCTCAGGAGGCTCCAAAGCTGAGCTGCTGCTTTATG 1560
DB 1501 TGAAGACTCGAAGCTCATCTCAGGAGGCTCCAAAGCTGAGCTGCTGCTTTATG 1560
QY 1561 ACCCGTGTGGGGAGAGAGAACTGAAAGTGTCTATCAAGTTCCGAGGCTCTGC 1620
DB 1561 ACCCGTGTGGGGAGAGAGAACTGAAAGTGTCTATCAAGTTCCGAGGCTCTGC 1620
QY 1621 ATCAGGTATGGGCTGAGCAGTGAAGGCTCCGAGTACCAAGAGCTCCACAGAGATG 1680
DB 1621 ATCAGGTATGGGCTGAGCAGTGAAGGCTCCGAGTACCAAGAGCTCCACAGAGATG 1680
QY 1681 ATACAGATGATTAACCTGCAAGAACCAAGATTTTAAAGGCGCAAAAATCTTTCT 1740
DB 1681 ATACAGATGATTAACCTGCAAGAACCAAGATTTTAAAGGCGCAAAAATCTTTCT 1740
QY 1741 GGGAGCTCAAAATTTGAAATGAAAAAACCAGACATAGATGTTTTATTTATTA 1800
DB 1741 GGGAGCTCAAAATTTGAAATGAAAAAACCAGACATAGATGTTTTATTTATTA 1800
QY 1801 TTAATTAAGAGTGGTACCATTTATCAATTATGTAAGGAGCATGACAGACCCAGCTT 1860
DB 1801 TTAATTAAGAGTGGTACCATTTATCAATTATGTAAGGAGCATGACAGACCCAGCTT 1860
QY 1861 TTGAGGAGTCTGGGGATGAGACTGAGGACGCCCATGAGAACCAAGTGCCTGCGC 1920
DB 1861 TTGAGGAGTCTGGGGATGAGACTGAGGACGCCCATGAGAACCAAGTGCCTGCGC 1920
QY 1921 CATGCTGTTTTCCCAAGATAGTTCTGAGAGGAAAGGCTGTGCTGACTCGCTG 1980
DB 1921 CATGCTGTTTTCCCAAGATAGTTCTGAGAGGAAAGGCTGTGCTGACTCGCTG 1980
```

QY 1881 TGTCCTGAGACACAGTGTGACCCGACGCCGCCCTGTAGTTCTTGGCTGAGT 2040  
 |||||  
 Db 1981 TGTCCTGAGACACAGTGTGACCCGACGCCGCCCTGTAGTTCTTGGCTGAGT 2040  
 |||||  
 QY 2041 GGTGTCTGTGAGACACCTGTGCTTCAACAAGAGCTGAGCCATTCTGCAATTCAGC 2100  
 |||||  
 Db 2041 GGTGTCTGTGAGACACCTGTGCTTCAACAAGAGCTGAGCCATTCTGCAATTCAGC 2100  
 |||||  
 QY 2101 TGAACATGAGGAAACAACCTGAAAGAGAGGCTCCCGGTGAGGAGCCTCTGCTGTG 2160  
 |||||  
 Db 2101 TGAACATGAGGAAACAACCTGAAAGAGAGGCTCCCGGTGAGGAGCCTCTGCTGTG 2160  
 |||||  
 QY 2161 CTGGCTTCCCATGACCAACCTCTCTGTGTAATATTACTCTTGAATCTGAGACAGATT 2220  
 |||||  
 Db 2161 CTGGCTTCCCATGACCAACCTCTCTGTGTAATATTACTCTTGAATCTGAGACAGATT 2220  
 |||||  
 QY 2221 GCGGGTTTATATAAATCTGTTTTTATCTGAGAAACAACGAGTTTGGAAATTAGTGTCTTT 2280  
 |||||  
 Db 2221 GCGGGTTTATATAAATCTGTTTTTATCTGAGAAACAACGAGTTTGGAAATTAGTGTCTTT 2280  
 |||||  
 QY 2281 TTTTCCCATCTCCGAGAGTGTCTCAAGTCAATCCACCGGCCCTCCGCTTGGGACAGGTT 2340  
 |||||  
 Db 2281 TTTTCCCATCTCCGAGAGTGTCTCAAGTCAATCCACCGGCCCTCCGCTTGGGACAGGTT 2340  
 |||||  
 QY 2341 AGTGTAACTCCGATGCCAGAGGCTGAGCCCTGACACAGTGGCTTCCGATCCCGGTGG 2400  
 |||||  
 Db 2341 AGTGTAACTCCGATGCCAGAGGCTTAGCCCTGACACAGTGGCTTCCGATCCCGGTGG 2400  
 |||||  
 QY 2401 GAAAAACGCTGCAACAGCGGGCTTGAAGTGGCTGTGTCTTCCACCGCTGACACAC 2460  
 |||||  
 Db 2401 GAAAAACGCTGCAACAGCGGGCTTGAAGTGGCTGTGTCTTCCACCGCTGACACAC 2460  
 |||||  
 QY 2461 CCACCTCCAGAGTGCAGTGTGAGGCAAGGGAGCTCAAGAGAGAGACAGAGGCTTGG 2520  
 |||||  
 Db 2461 CCACCTCCAGAGTGCAGTGTGAGGCAAGGGAGCTCAAGAGAGAGACAGAGGCTTGG 2520  
 |||||  
 QY 2521 CAAGACATCAAGACACCAACCAAGAGGCTGAGACCCAGAGCCCGGCTGTGTAACCA 2580  
 |||||  
 Db 2521 CAAGACATCAAGACACCAACCAAGAGGCTGTGAGACCCAGAGCCCGGCTGTGTAACCA 2580  
 |||||  
 QY 2581 GCAGGTGGCACTGAGACTCCCGCTCTCTGACAGTTCACAGCTCTTCAACAGAAACAAGG 2640  
 |||||  
 Db 2581 GCAGGTGGCACTGAGACTCCCGCTCTCTGACAGTTCACAGCTCTTCAACAGAAACAAGG 2640  
 |||||  
 QY 2641 CCTGTGCTCCGAGACCTTCTTCAAGACCTTCTTCAAGTGCACATTTGGGATCAGAAAT 2700  
 |||||  
 Db 2641 CCTGTGCTCCGAGACCTTCTTCAAGACCTTCTTCAAGTGCACATTTGGGATCAGAAAT 2700  
 |||||  
 QY 2701 GCAGCGAGAGTAGGACCCCTCCACAGGCTGTGACCTCGGCTGAGTAAAGTTAGTAGG 2760  
 |||||  
 Db 2701 GCAGCGAGAGTAGGACCCCTCCACAGGCTGTGACCTCGGCTGAGTAAAGTTAGTAGG 2760  
 |||||  
 QY 2761 CTTGTCTCTCGGGGCTGTGAAAGTGGACCATCAGTTGCTTGTGTAACCCCTCGAGACA 2820  
 |||||  
 Db 2761 CTTGTCTCTCGGGGCTGTGAAAGTGGACCATCAGTTGCTTGTGTAACCCCTCGAGACA 2820  
 |||||  
 QY 2821 AGCCGCGGACAGGTGTGTGTGAGACAGCTGTGCGGGGGGCCCCAGCTGCGCGGCTT 2880  
 |||||  
 Db 2821 AGCCGCGGACAGGTGTGTGTGAGACAGCTGTGCGGGGGGCCCCAGCTGCGCGGCTT 2880  
 |||||  
 QY 2881 CCAGCCCAACCCACAGCTGTGTGAGTCAAGGCCAATCTCCACGACTGTGTATCTGAGT 2940  
 |||||  
 Db 2881 CCAGCCCAACCCACAGCTGTGTGAGTCAAGGCCAATCTCCACGACTGTGTATCTGAGT 2940  
 |||||  
 QY 2941 AACGGCTTAAGAACTCTTCTCTGTGTGTTGAAAGAGTTGCGGTTGTCCAAATTCGTA 3000  
 |||||  
 Db 2941 AACGGCTTAAGAACTCTTCTCTGTGTGTTGAAAGAGTTGCGGTTGTCCAAATTCGTA 3000  
 |||||  
 QY 3001 ACATTTCATCTCATTATTTTAAAGAGTTTCTGACGCGGCCCAAGCGCCGCGGT 3060  
 |||||  
 Db 3001 ACATTTCATCTCATTATTTTAAAGAGTTTCTGACGCGGCCCAAGCGCCGCGGT 3060  
 |||||  
 QY 3061 GAGGCTGTGTGTGATGAGCTGTGAGGCCCGGGGCTTCCGCTGCGCTGCGCAGAGTGTCT 3120  
 |||||

Db 3061 GAGGCTGTGTGTGATGAGCTGTGAGGCCCGGGGCTTCCGCTGCGCAGAGTGTCT 3120  
 |||||  
 QY 3121 TCTGGGACACCATCTCTGTGCTTCAATTTGCACTGCAGTCTACAGAAAGCATCTACACA 3180  
 |||||  
 Db 3121 TCTGGGACACCATCTCTGTGCTTCAATTTGCACTGCAGTCTACAGAAAGCATCTACACA 3180  
 |||||  
 QY 3181 ATAAACCTTCTGAAAGCAGAAAAA 3213  
 |||||  
 Db 3181 ATAAACCTTCTGAAAGCAGAAAAA 3213  
 |||||  
 RESULT 2  
 ABS64608  
 ID ABS64608 standard; cDNA; 3230 BP.  
 XX  
 AC ABS64608;  
 XX  
 DT 15-NOV-2002 (first entry)  
 XX  
 DE Human cDNA probe selectively hybridising to foetal cell mRNA #23.  
 XX  
 KW Human; se; foetal liver myeloid cell; probe; erythroblast;  
 KW foetal abnormality; maternal blood; differential display;  
 KW chromosomal abnormality; single gene disorder; aneuploidy;  
 KW nucleotide triplet expansion disorder; trisomy 13; trisomy 21;  
 KW Klinefelter syndrome; spina bifida; sickle cell anaemia; thalassemia;  
 KW Marfan syndrome; Duchenne muscular dystrophy; cystic fibrosis;  
 KW Fragile X-syndrome; Friedreich's ataxia; myotonic dystrophy;  
 KW Huntington's disease; sex determination.  
 XX  
 OS Homo sapiens.  
 XX  
 PN W0200255985-A2.  
 XX  
 PD 18-JUL-2002.  
 XX  
 PF 01-NOV-2001; 2001MO-US045340.  
 XX  
 PR 15-NOV-2000; 2000US-0248882P.  
 XX  
 PA (HOPF ) ROCHE DIAGNOSTICS CORP.  
 XX  
 PI Schueler PA, Xu H, Folta L, Wu X, Sha Y, Nagy A, Mahoney WC;  
 XX  
 DR WPI; 2002-619108/66.  
 DR  
 DR P-PSDB; ABG79328.  
 XX  
 PT Detecting fetal cells in maternal blood sample, useful for diagnosing an  
 PT abnormality in a fetal cell, involves using specific nucleic acid probes  
 PT that hybridize to fetal cell associated RNAs.  
 XX  
 PS Claim 109; Page 169-170; 215pp; English.  
 XX  
 CC The invention relates to detecting foetal cells in maternal blood sample,  
 CC involving contacting sample with a first probe or performing expression  
 CC analysis on RNA/cDNA obtained from foetal liver myeloid cells relative to  
 CC that obtained from mature cells to identify RNA/cDNA species useful as  
 CC probe, contacting sample with the probe and identifying if the sample  
 CC comprises a cell that comprises mRNA that hybridises to the probe. The  
 CC probes are identified by differential display analysis using mature liver  
 CC cells and foetal liver myeloid cells of less than 22 weeks of gestation.  
 CC Also included are: (1) The probe sequences (appearing as ABS64586-  
 CC ABS64618), an isolated nucleic acid molecule having a sequence which is  
 CC at least 90% identical to the probe or its complement, or having a  
 CC nucleotide sequence identical to at least 15 consecutive nucleotide  
 CC residues of the probe, the encoded proteins from the probe or encoding a  
 CC naturally occurring allelic variant or fragment; (2) a non-mammalian host  
 CC cell containing the probe and (3) an antibody which selectively binds to  
 CC the probe encoded protein. The method is useful for detecting a foetal  
 CC cell (such as erythroblast or trophoblast) in a maternal blood sample,  
 CC which is useful for diagnosing a chromosomal abnormality, single gene  
 CC disorder or nucleotide triplet expansion in the gene, in a foetal cell.

CC For example an aneuploidy (trisomy 13, trisomy 21, or Klinefelter  
CC syndrome), spina bifida, sickle cell anaemia, a thalassemia, Marfan  
CC syndrome, Duchenne muscular dystrophy, cystic fibrosis, Fragile X-  
CC syndrome, Friedreich's ataxia, myotonic dystrophy or Huntington's  
CC disease. The probes are also useful for determination of the sex of a  
CC fetus and for detecting the presence of multiple foetuses at early  
CC stages of pregnancy. The present sequence is a foetal cell specific probe  
CC of the invention  
XX

Sequence 3230 BP; 782 A; 875 C; 897 G; 676 T; 0 U; 0 Other;

Query Match 99.3%; Score 3191.4; DB 6; Length 3230;

Best Local Similarity 99.8%; Pred. No. 0;

Matches 3206; Conservative 0; Mismatches 6; Indels 1; Gaps 1;

```
Qy 1 GAAAGGTTGGAGAGATGCGCCGCTTGGAGCGAGAGAGCTGGACATGAAGACTATT 60
Db 7 GAAAGGTTGGAGAGATGCGCCGCTTGGAGCGAGAGAGCTGGACATGAAGACTATT 66
Qy 61 ACTGTTGCTGAACGTGCGAGAGAGGCTCTTCTGAAGAGCTGAAAGCTGCTACCGGA 120
Db 67 ACTGTTGCTGAACGTGCGAGAGAGGCTCTTCTGAAGAGCTGAAAGCTGCTACCGGA 126
Qy 121 GGCTCTGTATGCTCTTACATCCAGACAGACAGAGACCCAGAGCTTCAAGTCAAGGCGG 180
Db 127 GGCTCTGTATGCTCTTACATCCAGACAGACAGAGACCCAGAGCTTCAAGTCAAGGCGG 186
Qy 181 AAGCAGCTGTTAACTCTTGTTCACCAAGCTTATGAAGTGTATGACCCCAACAGAGG 240
Db 187 AAGCAGCTGTTAACTCTTGTTCACCAAGCTTATGAAGTGTATGACCCCAACAGAGG 246
Qy 241 CCATCTATGATATATATGGAAGAGAGAGCTGAAATGGAAGATGGAGGTTGTGGA 300
Db 247 CCATCTATGATATATATGGAAGAGAGAGCTGAAATGGAAGATGGAGGTTGTGGA 306
Qy 301 GAGAGAGAACCCCTGCTGAATTCGAGAGAGTTTGAAGCGCTGACAGAGAGAGAG 360
Db 307 GAGAGAGAACCCCTGCTGAATTCGAGAGAGTTTGAAGCGCTGACAGAGAGAGAG 366
Qy 361 AGAGAGATGTCAGACAGACCAATCCCAAGGGAAGATCAGCGTTGGATGATGCCA 420
Db 367 AGAGAGATGTCAGACAGACCAATCCCAAGGGAAGATCAGCGTTGGATGATGCCA 426
Qy 421 CCGACCTTTTGAATCGTATGATGAGAGATGAAGATGTCGCGAGTGAAGCTTTCGG 480
Db 427 CCGACCTTTTGAATCGTATGATGAGAGATGAAGATGTCGCGAGTGAAGCTTTCGG 486
Qy 481 AGATTGAATTTAATAAATGCAATATCCAGTCCATTGAGGCAACCTTGAACAGCAG 540
Db 487 AGATTGAATTTAATAAATGCAATATCCAGTCCATTGAGGCAACCTTGAACAGCAG 546
Qy 541 ACAAGGCAATCTCTCTGGAAGCCTCTCAACCCAGAAATGGAATGAGAGAGTTTCA 600
Db 547 ACAAGGCAATCTCTCTGGAAGCCTCTCAACCCAGAAATGGAATGAGAGAGTTTCA 606
Qy 601 ACTTGGGCTCAGACGAGTAACCTTGGCAAGGAGATGGGAGATTGGAATTTGAGCTG 660
Db 607 ACTTGGGCTCAGACGAGTAACCTTGGCAAGGAGATGGGAGATTGGAATTTGAGCTG 666
Qy 661 GAGACCTTACAGGGGCTTTTGTGCTCAAGCTGTTCCGTAAATCTTCAACCAAGATGCT 720
Db 667 GAGACCTTACAGGGGCTTTTGTGCTCAAGCTGTTCCGTAAATCTTCAACCAAGATGCT 726
Qy 721 TTGTGACAAACAACTGCTCTGCAAGTTTATCCGTGAATCCGACCGGCTGACCA 780
Db 727 TTGTGACAAACAACTGCTCTGCAAGTTTATCCGTGAATCCGACCGGCTGACCA 786
Qy 781 CTGTCTAGCTCGGAACCTTGAACAAGAACCCGTGGCTACTGCAAGTGGCGATGGGGTA 840
Db 787 CTGTCTAGCTCGGAACCTTGAACAAGAACCCGTGGCTACTGCAAGTGGCGATGGGGTA 846
Qy 841 TCCAGTACGCATGAACATAGATCGTCCGAGACATTAACACAGCACTTCACTGTGG 900
Db 847 TCCAGTACGCATGAACATAGATCGTCCGAGACATTAACACAGCACTTCACTGTGG 906
```

```
Db 847 TCCAGTACGCATGAACATAGATCGTCCGAGACATTAACACAGCACTTCACTGTGG 906
Qy 901 CCTGTAGCTGGAAATCCCTCATCTCTTGGACATGATGATAGCAAAATTCAG 960
Db 907 CCTGTAGCTGGAAATCCCTCATCTCTTGGACATGATGATAGCAAAATTCAG 966
Qy 961 ATGACATGACATCGTGTGAAGAGATCCCTCAAGCAGGCTTCTTTGGGACGCTGTGG 1020
Db 967 ATGACATGACATCGTGTGAAGAGATCCCTCAAGCAGGCTTCTTTGGGACGCTGTGG 1026
Qy 1021 AGTACGAGCTGAGAGAGATCTCCAGGACAGAGCTTTTGGGTGACGTTCAGCGTTGG 1080
Db 1027 AGTACGAGCTGAGAGAGATCTCCAGGACAGAGCTTTTGGGTGACGTTCAGCGTTGG 1086
Qy 1081 GAGTTCACAGGAGCGTTTCTCAAAAGTCAACAGGAGCTCAAGGAGCTGACATCTTCT 1140
Db 1087 GAGTTCACAGGAGCGTTTCTCAAAAGTCAACAGGAGCTCAAGGAGCTGACATCTTCT 1146
Qy 1141 TCCCTATTCATCTGACGACAGCTTTCGCGAGGCGCATGTTCTATGCAACCGTGGGAC 1200
Db 1147 TCCCTATTCATCTGACGACAGCTTTCGCGAGGCGCATGTTCTATGCAACCGTGGGAC 1206
Qy 1201 CTCTAGTGTCTACTTTTGCATGCAACCGTCTGATCATCAACCATATCTCAGGGCTCAGA 1260
Db 1207 CTCTAGTGTGTACTTTTGCATGCAACCGTCTGATCATCAACCATATCTCAGGGCTCAGA 1266
Qy 1261 AAGAGAGGAATTTGAGAGAGAGAGAGAGGCGGCAAGTGTGTGAGAGAGAGC 1320
Db 1267 AAGAGAGGAATTTGAGAGAGAGAGAGAGGCGGCAAGTGTGTGAGAGAGAGC 1326
Qy 1321 AAGAGGCGAGATCCGCTGTCCGCTGATGCAAGAAATCTGTCCGAAGGATTAATTGAGGCGAG 1380
Db 1327 AAGAGGCGGAGATCCGCTGTCCGCTGATGCAAGAAATCTGTCCGAAGGATTAATTGAGGCGAG 1386
Qy 1381 AAGAGTCCAGAAATGGGCTCTCATGTCATATGCTGTGTAAGGAAATTTGCAATGACA 1440
Db 1387 AAGAGTCCAGAAATGGGCTCTCATGTCATATGCTGTGTAAGGAAATTTGCAATGACA 1446
Qy 1441 AAGAGCAGAGAGAGAGAGAGAGAGAGAGTGAAGTGAAGTGTGAGGAGAGTGTGAG 1500
Db 1447 AAGAGCAGAGAGAGAGAGAGAGAGAGAGTGAAGTGAAGTGTGAGGAGAGTGTGAG 1506
Qy 1501 TGAAGGACTCGAAGGCTCATCTCAGAGAGGCTTCAAGAGGCTGAGGCTTCTTATG 1560
Db 1507 TGAAGGACTCGAAGGCTCATCTCAGAGAGGCTTCAAGAGGCTGAGGCTTCTTATG 1566
Qy 1561 ACCCGTGTGTGGGAGAGAGAGAACTGAAGATGCTCTATCATGTTCCGGGCGCTCTGC 1620
Db 1567 ACCCGTGTGTGGGAGAGAGAGAACTGAAGATGCTCTATCATGTTCCGGGCGCTCTGC 1626
Qy 1621 ATCAGGTGATGAGTGTGCAAGTGAAGGCGCTCCGAGTACCAAGGAGTCCACAGAGATGG 1680
Db 1627 ATCAGGTGATGAGTGTGCAAGTGAAGGCGCTCCGAGTACCAAGGAGTCCACAGAGATGG 1686
Qy 1681 ATACAGATGATTAACCTGCAAGAACTCAGATTTTAAAGGCGCGCAAAAATCTTTTCT 1740
Db 1687 ATACAGATGATTAACCTGCAAGAACTCAGATTTTAAAGGCGCGCAAAAATCTTTTCT 1746
Qy 1741 GGGAGTCTCAAAATTTGGAATGAAAAAACCAGACATCAGATGTTTTTATTTATATTA 1800
Db 1747 GGGAGTCTCAAAATTTGGAATGAAAAAACCAGACATCAGATGTTTTTATTTATATTA 1806
Qy 1801 TTAATTAATAAGTGTATCAATTAATTAATTAAGTGAAGAGATGACAGACCCAGCTT 1860
Db 1807 TTAATTAATAAGTGTATCAATTAATTAATTAAGTGAAGAGATGACAGACCCAGCTT 1866
Qy 1861 TTGAGGGTCTGGGGGTAGAGCTGAGAGAGCCCACTGGGAACAGAGCTGACCTGGGC 1920
Db 1867 TTGAGGGTCTGGGGGTAGAGCTGAGAGAGCCCACTGGGAACAGAGCTGACCTGGGC 1926
Qy 1921 CATGGCTGTTTTCCAGAGATCATGTTCTGAGGGAGAGGGCTGTGGCCTGACTCCGCTG 1980
Db 1927 CATGGCTGTTTTCCAGAGATCATGTTCTGAGGGAGAGGGCTGTGGCCTGACTCCGCTG 1986
```

Oy 1981 TGTCCGAGCA CAGGTGCTGACCCGAGCCGCGCCCTGTAGTTCTTGGCTGGGTCTGGA 2040  
 Db 1987 TGTCCGAGCA CAGGTGCTGACCCGAGCCGCGCCCTGTAGTTCTTGGCTGGGTCTGGA 2046  
 Oy 2041 GGTGTCTGTGAGCA CCGTGCCTCTACCA CAGAGCTGAGCACTTCTGAGTCCAGC 2100  
 Db 2047 GGTGTCTGTGAGCA CCGTGCCTCTACCA CAGAGCTGAGCACTTCTGAGTCCAGC 2106  
 Oy 2101 TGAACATGGGAAACA CTTGAAAGAGCAGGCTCTCCGGTCA GGGAGCTCTGTGTG 2160  
 Db 2107 TGAACATGGGAAACA CTTGAAAGAGCAGGCTCTCCGGTCA GGGAGCTCTGTGTG 2166  
 Oy 2161 CTGGCTTCCATGAC CACCTCTCTGCTGAAATTTA CTGCTTGAATCTGGAGAGATT 2220  
 Db 2167 CTGGCTTCCATGAC CACCTCTCTGCTGAAATTTA CTGCTTGAATCTGGAGAGATT 2226  
 Oy 2221 GCGGGTTTATAAATCTGCTTTTATCTGAGAA CAAACGGGTTTGGAAATTAGTCTCTT 2280  
 Db 2227 GCGGGTTTATAAATCTGCTTTTATCTGAGAA CAAACGGGTTTGGAAATTAGTCTCTT 2286  
 Oy 2281 TTTCCCACTCCAGAGCTGCTCAAGTCA TTCACCGGCCCCCTGCGCTTGGAGACGGT 2340  
 Db 2287 TTTCCCACTCCAGAGCTGCTCAAGTCA TTCACCGGCCCCCTGCGCTTGGAGACGGT 2346  
 Oy 2341 AGTGAATCTCCGAGATCCAGGGGCTTGA GCGCTGACACAGGTGGCTTCCGGTATCCCGGTG 2400  
 Db 2347 AGTGAATCTCCGAGATCCAGGGGCTTGA GCGCTGACACAGGTGGCTTCCGGTATCCCGGTG 2406  
 Oy 2401 GAAAAAGCCCTGCA CAGCGGGCTTGA GCTGAGCTGCTGTGCTCTCA CCGGCTGACAC 2460  
 Db 2407 GAAAAAGCCCTGCA CAGCGGGCTTGA GCTGAGCTGCTGTGCTCTCA CCGGCTGACAC 2466  
 Oy 2461 CCACTTCCAGAGTGA GTCCTGGGCAAGGGCA GCTCAAGAGACAGACCAAGCGCTTGG 2520  
 Db 2467 CCACCTCCAGAGTGA GTCCTGGGCAAGGGCA GCTCAAGAGACAGACCAAGCGCTTGG 2526  
 Oy 2521 CAAGCATATGACA CACCCCAACCCAAAGGCTGAG CCCCAGGCGCGGCTGTGATCCCA 2580  
 Db 2527 CAAGCATATGACA CACCCCAACCCAAAGGCTGAG CCCCAGGCGCGGCTGTGATCCCA 2586  
 Oy 2581 GCAGGTGACATGCA GCTCCCGCTCTCTGACAGTTCAGAGTCTCTTCA CAGAAACACAGGG 2640  
 Db 2587 GCAGGTGACATGCA GCTCCCGCTCTCTGACAGTTCAGAGTCTCTTCA CAGAAACACAGGG 2646  
 Oy 2641 CCTGTGCTCGGAG CTTCTTTCAGACCCTTCTTCCAGTGCCTCA CTTGGAGTGCAGAT 2700  
 Db 2647 CCTGTGCTCGGAG CTTCTTTCAGACCCTTCTTCCAGTGCCTCA CTTGGAGTGCAGAT 2706  
 Oy 2701 GCAGGGAGCTAGGAC CCCCCTCCACGGCTGGA CTTGGGCTGCAATAAGTTAAGTTAGTGA 2760  
 Db 2707 GCAGGGAGCTAGGAC CCCCCTCCACGGCTGGA CTTGGGCTGCAATAAGTTAAGTTAGTGA 2766  
 Oy 2761 CCTGTCTCTCGG GGCCTGGAAGTGCAGCCATCAGTTCTCTTGTGTA CCCCCTCGAGCA 2820  
 Db 2767 CCTGTCTCTCGG GGCCTGGAAGTGCAGCCATCAGTTCTCTTGTGTA CCCCCTCGAGCA 2826  
 Oy 2821 AGCCGCGACAGTGTG GCTGAGACAGCTGAGCGGAGGAGGCCCAAGCTGCGCGGCT 2880  
 Db 2827 AGCCGCGACAGTGTG GCTGAGACAGCTGAGCGGAGGAGGCCCAAGCTGCGCGGCT 2886  
 Oy 2881 CCAGCCCAACCAAG CTTGCTGTAAGTCAAGCCCACTCCCAAGCACTGTATCTGAGT 2940  
 Db 2887 CCAGCCCAACCAAG CTTGCTGTAAGTCAAGCCCACTCCCAAGCACTGTATCTGAGT 2946  
 Oy 2941 AACGGCTAAGAACTCTTCTCTGTGTTTGA AAAAGAGTTCGGGTGTCTCAATTCTGTA 3000  
 Db 2947 AACGGCTAAGAACTCTTCTCTGTGTTTGA AAAAGAGTTCGGGTGTCTCAATTCTGTA 3006  
 Oy 3001 ACATTCAATCTCAATTTTAAAAAGGTTTCTGAGCGGCCCAAGCGCCGAGCGCGGT 3060  
 Db 3007 ACATTCAATCTCAATTTTAAAAAGGTTTCTGAGCGGCCCAAGCGCCGAGCGCGGT 3065

Oy 3061 GAGCGTGTGTGATGAGAGCTGAGCCCGGAGCTTCCCGTGCCTCTGCGCAGAGTGTCT 3120  
 Db 3066 GAGCGTGTGTGATGAGAGCTGAGCCCGGAGCTTCCCGTGCCTCTGCGCAGAGTGTCT 3125  
 Oy 3121 TCTGGGACCCATCTCTGCGTTTCATTGTCAGTGA CTGTACAGAAAGCACTACCA 3180  
 Db 3126 TCTGGGACCCATCTCTGCGTTTCATTGTCAGTGA CTGTACAGAAAGCACTACCA 3185  
 Oy 3181 ATAAACCTTTCTGTAAGACGAAAAA 3213  
 Db 3186 ATAAACCTTTCTGTAAGACGAAAAA 3218  
 RESULT 3  
 AAH14482  
 ID AAH14482 standard; cDNA; 3199 BP.  
 AC AAH14482;  
 XX  
 XX  
 DT 26-JUN-2001 (first entry)  
 XX  
 DE Human cDNA sequence SEQ ID NO:11986.  
 XX  
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
 XX Homo sapiens.  
 OS  
 PN EP1074617-A2.  
 PD  
 XX 07-FEB-2001.  
 PF 28-JUL-2000; 2000EP-00116126.  
 XX  
 PR 29-JUL-1999; 99JP-00248036.  
 PR 27-AUG-1999; 99JP-00300253.  
 PR 11-JAN-2000; 2000JP-00118776.  
 PR 02-MAY-2000; 2000JP-00183767.  
 PR 09-JUN-2000; 2000JP-00241899.  
 XX  
 PA (HELI-) HELIX RES INST.  
 XX  
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 PI Ishii S, Sugiyama T, Wakamatsu A, Negai K, Otsuki T;  
 XX WPI; 2001-318749/34.  
 DR  
 XX  
 XX  
 PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.  
 PT  
 PT  
 XX  
 XX  
 XX  
 Claim 8; SEQ ID NO 11986; 2537pp + Sequence listing; English.  
 XX  
 CC The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification, where a primer set comprises:  
 CC (a) an oligo-dT primer and an oligonucleotide complementary to the  
 CC complementary strand of a polynucleotide which comprises one of the 5602  
 CC nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence and 3'-end sequence is selected from those defined in the  
 CC specification. The primer sets can be used in antisense therapy and in  
 CC gene therapy. The primers are useful for synthesizing polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialized methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893  
 CC represent human amino acid sequences; and AAH13629 to AAH13632 represent

CC oligonucleotides, all of which are used in the exemplification of the  
CC present invention  
XX  
SQ Sequence 3199 BP; 759 A; 875 C; 890 G; 675 T; 0 U; 0 Other;

Query Match 98.7%; Score 3171.6; DB 4; Length 3199;  
Best Local Similarity 99.7%; Pred. No. 0;  
Matches 3188; Conservative 0; Mismatches 9; Indels 1; Gaps 1;

```
QY 2 AAAGTTGGAGATGGCGAGCGCTTGAGCGAGGAGCTGGACAATGAACATATTA 61
   |||
Db 3 AAAGTTGGAGATGGCGAGCGCTTGAGCGAGGAGCTGGACAATGAACATATTA 62
QY 62 CTCGTGCTGAAGCGTGCAGAGGAGCGCTTCTGGAAGCTGAAAAGCTCTACCGGAG 121
   |||
Db 63 CTCGTGCTGAAGCGTGCAGAGGAGCGCTTCTGGAAGCTGAAAAGCTCTACCGGAG 122
QY 122 GCTCTGATGCTCTACCATCCAGACACAGAGCCAGAGCTCAAGTCAAGCGGGA 181
   |||
Db 123 GCTCTGATGCTCTACCATCCAGACACAGAGCCAGAGCTCAAGTCAAGCGGGA 182
QY 182 ACAGCTGTTAACCTTGTCACAGCGCTTATGAAGTCTAGTGAACCCCAACAGGCG 241
   |||
Db 183 ACAGCTGTTAACCTTGTCACAGCGCTTATGAAGTCTAGTGAACCCCAACAGGCG 242
QY 242 CATCTATGATATATATGGAAGAGAGAGCTGGAATGGAAGATGGAGGTTGTGGAAG 301
   |||
Db 243 CATCTATGATATATATGGAAGAGAGAGCTGGAATGGAAGATGGAGGTTGTGGAAG 302
QY 302 GAGAGAAACCCCTGCTGTAATTTGAGAGAGGTTGAGCGGCTGACAGAGAGAGAGA 361
   |||
Db 303 GAGAGAAACCCCTGCTGTAATTTGAGAGAGGTTGAGCGGCTGACAGAGAGAGAGA 362
QY 362 GAGAGATTCAGAGCGAAACCAATCCCAAGGGAAGATCAGGTTGAGTATAGTGCAC 421
   |||
Db 363 GAGAGATTCAGAGCGAAACCAATCCCAAGGGAAGATCAGGTTGAGTATAGTGCAC 422
QY 422 CGACCTTTTGTATCGTATGATGAGAGATGGAAGTGTCCGCGAGTACTTTCGCA 481
   |||
Db 423 CGACCTTTTGTATCGTATGATGAGAGATGGAAGTGTCCGCGAGTACTTTCGCA 482
QY 482 GATTGAAATTAATAATGACATATCCAGTCCATTGAGGCACCTTGAACAGACAGA 541
   |||
Db 483 GATTGAAATTAATAATGACATATCCAGTCCATTGAGGCACCTTGAACAGACAGA 542
QY 542 CACAGCATCTCTCTGGAAGCCTCTCAACCCAGAAATGGAATGAGGAGGTTCCATTAA 601
   |||
Db 543 CACAGCATCTCTCTGGAAGCCTCTCAACCCAGAAATGGAATGAGGAGGTTCCATTAA 602
QY 602 CTTTGCGCTCAGAGAGTAACTTCGCGAAAGGATGGGAGAGTGGAAATTTGAGCTGG 661
   |||
Db 603 CTTTGCGCTCAGAGAGTAACTTCGCGAAAGGATGGGAGAGTGGAAATTTGAGCTGG 662
QY 662 AGACCTTACAGGGGCTTTTGTCCGTCTCAGAGCTGTTCCGTAATCTCAACCAAGATGCTT 721
   |||
Db 663 AGACCTTACAGGGGCTTTTGTCCGTCTCAGAGCTGTTCCGTAATCTCAACCAAGATGCTT 722
QY 722 TGTGACAAACAACTGTGCTCTGACAGTTCATCCCGTGAATCGAACCCGCGCTGACAC 781
   |||
Db 723 TGTGACAAACAACTGTGCTCTGACAGTTCATCCCGTGAATCGAACCCGCGCTGACAC 782
QY 782 TGTCTAGCTCGAACCCTAGACAAACACCGTGGGCTACCTGCAAGTGGGAT 841
   |||
Db 783 TGTCTAGCTCGAACCCTAGACAAACACCGTGGGCTACCTGCAAGTGGGAT 842
QY 842 CCACTCAGCATGAACTATGACATGCTGCGAGACATTAACCAAGCCACTTCACTGTGCG 901
   |||
Db 843 CCACTCAGCATGAACTATGACATGCTGCGAGACATTAACCAAGCCACTTCACTGTGCG 902
QY 902 CCACTCAGCATGAACTATGACATGCTGCGAGACATTAACCAAGCCACTTCACTGTGCG 961
   |||
Db 903 CCACTCAGCATGAACTATGACATGCTGCGAGACATTAACCAAGCCACTTCACTGTGCG 962
```

```
QY 962 TGAAGATCAGACTCGTGTGAAAAGATCCCTCAAGACAGGCTTCTTTGGACGGTGTGGA 1021
   |||
Db 963 TGAAGATCAGACTCGTGTGAAAAGATCCCTCAAGACAGGCTTCTTTGGACGGTGTGGA 1022
QY 1022 GTTACGAGCTGAGAGGAATCTCCAGGACACAGCGTTTGGGTGACGCTGTACGCTTGG 1081
   |||
Db 1023 GTTACGAGCTGAGAGGAATCTCCAGGACACAGCGTTTGGGTGACGCTGTACGCTTGG 1082
QY 1082 AGTTCCAGAGGGGCTTCTCTCAAAAGTCAAGCTCAACAGGGCCAGTACATCTCTT 1141
   |||
Db 1083 AGTTCCAGAGGGGCTTCTCTCAAAAGTCAAGCTCAACAGGGCCAGTACATCTCTT 1142
QY 1142 CCTATATCACTTGAACGACCACTTTCGCGACGCGCATGTTCTATGCCCGTGGGCGC 1201
   |||
Db 1143 CCTATATCACTTGAACGACCACTTTCGCGACGCGCATGTTCTATGCCCGTGGGCGC 1202
QY 1202 TCTAGTGTCTACTTTTGGCATGACCCGTGTGATCATCAAAACCATACCTTACGGGCTCAGA 1261
   |||
Db 1203 TCTAGTGTCTACTTTTGGCATGACCCGTGTGATCATCAAAACCATACCTTACGGGCTCAGA 1262
QY 1262 AGAAGAAATTTGAGAAACAGAGGGAAGCGCCGCAACGATGTGCTCAGAAAGAGA 1321
   |||
Db 1263 AGAAGAAATTTGAGAAACAGAGGGAAGCGCCGCAACGATGTGCTCAGAAAGAGA 1322
QY 1322 AGAGCGAAGTCCGCTGTCGCGCTGATGCAAGGAATCTGTCCGAGGATATTTAGGCGA 1381
   |||
Db 1323 AGAGCGAAGTCCGCTGTCGCGCTGATGCAAGGAATCTGTCCGAGGATATTTAGGCGA 1382
QY 1382 AGAGTCCAAATATGGGCTCATCATGTGCAATGCTGTGTCGGGAAGTTTGTCAATGACA 1441
   |||
Db 1383 AGAGTCCAAATATGGGCTCATCATGTGCAATGCTGTGTCGGGAAGTTTGTCAATGACA 1442
QY 1442 GAGCAGGAAGAGCAGAAAGGTGAAGGTGATGACGTAATGAGCTGAGCCCTGACAGTGCCTGAT 1501
   |||
Db 1443 GAGCAGGAAGAGCAGAAAGGTGAAGGTGATGACGTAATGAGCTGAGCCCTGACAGTGCCTGAT 1502
QY 1502 GAAAGACTGGAAGCTCATCTTCAACGAGGCGTCCAAAGCTGAGGCTGCTGCTTTTATGA 1561
   |||
Db 1503 GAAAGACTGGAAGCTCATCTTCAACGAGGCGTCCAAAGCTGAGGCTGCTGCTTTTATGA 1562
QY 1562 CCCGTGTGCGGGGGAAGAAAGAACTTGAAATGTCTATACGATTCGCGGGGCTCTGGA 1621
   |||
Db 1563 CCCGTGTGCGGGGGAAGAAAGAACTTGAAATGTCTATACGATTCGCGGGGCTCTGGA 1622
QY 1622 TCAGGTGATGTGCTGACAGTGAAGGCTCCGGAATCAACAAAGAGTCCCAAGATGCA 1681
   |||
Db 1623 TCAGGTGATGTGCTGACAGTGAAGGCTCCGGAATCAACAAAGAGTCCCAAGATGCA 1682
QY 1682 TACAGATGATTAACCTGCCAAGAACCAATTTTAAAGGCGCAAAAATCTTTTCTGT 1741
   |||
Db 1683 TACAGATGATTAACCTGCCAAGAACCAATTTTAAAGGCGCAAAAATCTTTTCTGT 1742
QY 1742 GGAATCTCAAAATTTGGAATGAAAAACCCAGACATCAGATGTTTATTTATATATAT 1801
   |||
Db 1743 GGAATCTCAAAATTTGGAATGAAAAACCCAGACATCAGATGTTTATTTATATATAT 1802
QY 1802 TATTAATGAAGGTGATACCATTAATATATGAAAGGACATGACACACCCAGCTT 1861
   |||
Db 1803 TATTAATGAAGGTGATACCATTAATATATGAAAGGACATGACACACCCAGCTT 1862
QY 1862 TGAAGGTGCTGAGGAGTGAAGCTGAGGACGCCCATGAGAACCAAGCTGAGCTGCGCC 1921
   |||
Db 1863 TGAAGGTGCTGAGGAGTGAAGCTGAGGACGCCCATGAGAACCAAGCTGAGCTGCGCC 1922
QY 1922 ATGAGCTGTTTCCCAAGATCAAGTTCCTGGAAGGAAGGGCTTGGCCCTGACCTCGCTGT 1981
   |||
Db 1923 ATGAGCTGTTTCCCAAGATCAAGTTCCTGGAAGGAAGGGCTTGGCCCTGACCTCGCTGT 1982
QY 1982 GTTCCGAGACACGCTGACAGCGACGCCGCGCTGTATGATTTCTTGGCTGGGCTGTGAG 2041
   |||
Db 1983 GTTCCGAGACACGCTGACAGCGACGCCGCGCTGTATGATTTCTTGGCTGGGCTGTGAG 2042
QY 2042 GTTCTGTGAGACACCTGCGCTTCAACACAGAGGCTGAGCACTTCTGCACTGCACTGCT 2101
```

Db 2043 GTGTCGTGAGACACCTGCCCCACACAGAGAGTGAGGCACTTTCGACGTCCAACT 2102  
Qy 2102 GAACATGGGAAACAACTGAAAAAGAGAGAGCCCTCCGGTCAGAGGACCTCTGTGTC 2161  
Db 2103 GAACATGGGAAACAACTGAAAAAGAGAGAGCCCTCCGGTCAGAGGACCTCTGTGTC 2162  
Qy 2162 TGGCTTCCCATGACCACTCCCTCGCTGAAATATTACTGCTGAAATCTGAGAGAGATTG 2221  
Db 2163 TGGCTTCCCATGACCACTCCCTCGCTGAAATATTACTGCTGAAATCTGAGAGAGATTG 2222  
Qy 2222 CGGGTTTATAAAACCTGCTTTTATCTGAGAAACAAACGGGTTTGAATAATTAGTCGCTTTT 2281  
Db 2223 CGGGTTTATAAAACCTGCTTTTATCTGAGAAACAAACGGGTTTGAATAATTAGTCGCTTTT 2282  
Qy 2282 TTCCCATCTCCAGAGCTGCTCAAGTATTCACCTGCCCCCTCGGCTTGGAGACAGGTA 2341  
Db 2283 TTCCCATCTCCAGAGCTGCTCAAGTATTCACCTGCCCCCTCGGCTTGGAGACAGGTA 2342  
Qy 2342 GTGTAACTCCCGATCCGAGGCTTAGCCCTGACACAGGTGGCTTCCGATCCCGGTGGG 2401  
Db 2343 GTGTAACTCCCGATCCGAGGCTTAGCCCTGACACAGGTGGCTTCCGATCCCGGTGGG 2402  
Qy 2402 AAAACGCCCTGCAACAGCGGGCTTGAAGTGGCTGTGCTCCCTCAACCGCTGACACACC 2461  
Db 2403 AAAACGCCCTGCAACAGCGGGCTTGAAGTGGCTGTGCTCCCTCAACCGCTGACACACC 2462  
Qy 2462 CACCTCCAGAGTGAAGTCTGGGACAGGGCAGCTCAAGAGAGACAGAGCAGGCGTTGGC 2521  
Db 2463 CACCTCCAGAGTGAAGTCTGGGACAGGGCAGCTCAAGAGAGACAGAGCAGGCGTTGGC 2522  
Qy 2522 AAGCATGACACACCCCAAGGCGTGAAGCCCAAGGCGCGGCGGTGATCCAG 2581  
Db 2523 AAGCATGACACACCCCAAGGCGTGAAGCCCAAGGCGCGGCGGTGATCCAG 2582  
Qy 2582 CAGGTGACATGACAGCTCCCGCTCTCTGACAGTCCAGGCTCTCAAGAAACACAGAGGC 2641  
Db 2583 CAGGTGACATGACAGCTCCCGCTCTCTGACAGTCCAGGCTCTCAAGAAACACAGAGGC 2642  
Qy 2642 CTGTGCTCCGAGCCTTCTTCAAGCCTTCTCTCAAGTCCCACTTGGAGTGAAGATG 2701  
Db 2643 CTGTGCTCCGAGCCTTCTTCAAGCCTTCTCTCAAGTCCCACTTGGAGTGAAGATG 2702  
Qy 2702 CAGGGAAGTGAAGCCCTCCAGAGCCTTGAAGCTCCGCTCAAGTAAAGTTACGTGAGGC 2761  
Db 2703 CAGGGAAGTGAAGCCCTCCAGAGCCTTGAAGCTCCGCTCAAGTAAAGTTACGTGAGGC 2762  
Qy 2762 CTGTCTCTCGGGGCTTGAAGTGAAGTCAAGTTCCTTGTGCTGACCCCTCGAGAGCA 2821  
Db 2763 CTGTCTCTCGGGGCTTGAAGTGAAGTCAAGTTCCTTGTGCTGACCCCTCGAGAGCA 2822  
Qy 2822 GCGCCGACAGAGTGTGCTGAGACAGTGGCGCGGGGGGCCCAAGCTGCGCGGCTTC 2881  
Db 2823 GCGCCGACAGAGTGTGCTGAGACAGTGGCGGTGGGGGCCCAAGCTGCGCGGCTTC 2882  
Qy 2882 CAGCCACCCACAGAGTGTGCTGAGAGTCAAGTCCCTCCCTCCACAGACGTGATCTGAATA 2941  
Db 2883 CAGCCACCCACAGAGTGTGCTGAGAGTCAAGTCCCTCCCTCCACAGACGTGATCTGAATA 2942  
Qy 2942 ACGGCTAAGAACCTCTCTCTCTGATTTGAAAAGCAGTTCGGGTTGCAAACTCTGTA 3001  
Db 2943 ACGGCTAAGAACCTCTCTCTCTGATTTGAAAAGCAGTTCGGGTTGCAAACTCTGTA 3002  
Qy 3002 CATTATCTCAATTTTAAAAAGGTTCTCTGAAGGCCCAAGCGGCCAGCGCGGTG 3061  
Db 3003 CATTATCTCAATTTTAAAAAGGTTCTCTGAAGGCCCAAGCGGCCAGCGCGGTG 3062  
Qy 3062 AGCGTCGTTGACAGTGAAGCTGGGCGCGGCTTCCCTGCGCTGAGCCAGAGTGTGTT 3121  
Db 3062 AGCGTCGTTGACAGTGAAGCTGGGCGCGGCTTCCCTGCGCTGAGCCAGAGTGTGTT 3121  
Qy 3122 CTGGGACACCATCTCTGCGTTTCAATTTGACAGTGAAGTCAAGAGGCACTCAACAA 3181

Db 3122 CTGGGACACCATCTCTGCGTTTCAATTTGACAGTGAAGTCAAGAGGCACTCAACAA 3181  
Qy 3182 TAAACCTTTCTGAAAC 3199  
Db 3182 TAAACCTTTCTGAAAC 3199  
RESULT 4  
AACT7352  
ID AACT7352 standard; cDNA; 2991 BP.  
XX  
AC AACT7352;  
XX  
DT 08-FEB-2001 (first entry)  
XX  
DE Human ORF2907 polynucleotide sequence SEQ ID NO:5813.  
XX  
KW Human; open reading frame; ORF; detection; cytosolic; hepatotropic;  
KW vulnery; antiparkinsonian; nootropic; neuroprotective;  
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;  
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;  
KW antiviral; antibacterial; antifungal; antineumatic; antihypoid;  
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;  
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
KW cholesterol ester storage; systemic lupus erythematosus; infection;  
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;  
KW thrombosis; contraceptive; se.  
XX  
OS Homo sapiens.  
XX  
PN MO200058473-A2.  
XX  
PD 05-OCT-2000.  
XX  
PF 31-MAR-2000; 2000MO-US008621.  
XX  
PR 31-MAR-1999; 99US-0127607P.  
XX  
PR 02-APR-1999; 99US-0127636P.  
XX  
PR 05-APR-1999; 99US-0127728P.  
XX  
PR 30-MAR-2000; 2000US-00540763.  
XX  
PA (CURA-) CURAGEN CORP.  
XX  
PI Shimkets RA, Leach M;  
XX  
DR WPI; 2000-602362/57.  
XX  
DR P-PSDB; AAB43143.  
XX  
PT Novel nucleic acids and peptides derived from open reading frame X,  
PT useful for treating e.g. cancers, proliferative disorders,  
PT neurodegenerative disorders and cardiovascular disease.  
XX  
PS Claim 5; Page 4978-4980; 5507P; English.  
XX  
CC AAC7446 to AACT7606 encode the proteins given in AAB40237 to AAB43397,  
CC which represent the human ORF open reading frames 1 to 3161. The ORF  
CC sequences have activities such as: cytosolic; hepatotropic; vulnery;  
CC antiparkinsonian; nootropic; neuroprotective; osteopathic;  
CC anticonvulsant; antiparkinsonian; nootropic; immunosuppressant; cardiant;  
CC candidant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;  
CC dermatological; immunosuppressive; antineumatic; antihypoid; antianaemic;  
CC antiviral; antifungal; antineumatic; antihypoid; antianaemic. The  
CC sequences can be used for determining the presence of or predisposition  
CC to, or preventing or treating pathological conditions associated with an  
CC ORF-associated disorder. The nucleic acids can be used to express ORF  
CC proteins in gene therapy vectors. The proteins and nucleic acids may be  
CC used to treat cancers, proliferative disorders, neurodegenerative  
CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,  
CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester



CC storage, systemic lupus erythematosus, severe combined immunodeficiency  
CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune  
CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and  
CC cartilage degeneration, nocturnal haemoglobinuria, anti-inflammatory diseases; to  
CC enhance coagulation; to inhibit thrombosis; and as a contraceptive

XX Sequence 2991 BP; 700 A; 817 C; 818 G; 655 T; 0 U; 1 Other;

Query Match 76.4%; Score 2456; DB 3; Length 2991;  
Best Local Similarity 99.7%; Pred. No. 0;  
Matches 2481; Conservative 0; Mismatches 5; Indels 2; Gaps 2;

QY 716 ATGCTTTGTGACAAACAACCTGCTGCTGCAAGTTTTCATCCCGTGAATCCGACCGGCT 775  
DB 505 AAGCTTTGTGACAAACAACCTGCTGCTGCAAGTTTTCATCCCGTGAATCCGACCGGCT 564  
QY 776 GACCACTGCTCTAGCTGCTGGAACCTGACCAAGAACACCTGCGGCTACCTGCAAGTGGCATG 835  
DB 565 GACCACTGCTCTAGCTGCTGGAACCTGACCAAGAACACCTGCGGCTACCTGCAAGTGGCATG 624  
QY 836 GGGTATCCAGTCCAGCCATGAACTAGCATGCTCCGAGACATGAAACGACCTTAC 895  
DB 625 GGGTATCCAGTCCAGCCATGAACTAGCATGCTCCGAGACATGAAACGACCTTAC 684  
QY 896 TGTGGCCCTGCACTGGGAAATCCCTCACTCTTTGCACTGATCAGTATCAGCAAAAT 955  
DB 685 TGTGGCCCTGCACTGGGAAATCCCTCACTCTTTGCACTGATCAGTATCAGCAAAAT 744  
QY 956 CCAAGATGACGATCAGACTGCTGTGAAAGATCCCT-CAAGCAGGCTTCTTTGGACGG 1014  
DB 745 CCAAGATGACGATCAGACTGCTGTGAAAGATCCCTGCAAGACAGGCTTCTTTGGACGG 804  
QY 1015 TGTGTGAGTACGAGAGTGAAGAGAGATCTCAAGGACAGCGTTTGGTGGAGCTGCA 1074  
DB 805 TGTGTGAGTACGAGAGTGAAGAGAGATCTCAAGGACAGCGTTTGGTGGAGCTGCA 864  
QY 1075 GCGTTGAGATTCACAGAGCGGTTTCTCTCAAGTCAAGCTCAACAGGCGCATGACAT 1134  
DB 865 GCGTTGAGATTCACAGAGCGGTTTCTCTCAAGTCAAGCTCAACAGGCGCATGACAT 924  
QY 1135 ACTTCTTCCCTATTTCACTTGAACGAGACAGCTTCCGCAAGCGCATGTTTATGACACCG 1194  
DB 925 ACTTCTTCCCTATTTCACTTGAACGAGACAGCTTCCGCAAGCGCATGTTTATGACACCG 984  
QY 1195 TGGGGCTCTAGTGTCTACTTTGCTCATGCAACGCTGATCATCAACCATACCTCAGG 1254  
DB 985 TGGGGCTCTAGTGTCTACTTTGCTCATGCAACGCTGATCATCAACCATACCTCAGG 1044  
QY 1255 CTCAGAAAGAGAGAAATTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1314  
DB 1045 CTCAGAAAGAGAGAAATTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1104  
QY 1315 AGAAGCAAG 1374  
DB 1105 AGAAGCAAG 1164  
QY 1375 AGGCAAGAGAGTCCAGATGGGCTCATCTGCTCATGCTGCTGGTACGGAGATTGTCA 1434  
DB 1165 AGGCAAGAGAGTCCAGATGGGCTCATCTGCTCATGCTGCTGGTACGGAGATTGTCA 1224  
QY 1435 ATGACAAAG 1494  
DB 1225 ATGACAAAG 1284  
QY 1495 GCGTGGAGAGAGAGTCAAGTCACTCTCAACGAGGCTCCAGGCTGGGCTGCTGGCT 1554  
DB 1285 GCGTGGAGAGAGTCAAGTCACTCTCAACGAGGCTCCAGGCTGGGCTGCTGGCT 1344  
QY 1555 TTTATGACCCGTGTGTGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1614  
DB 1345 TTTATGACCCGTGTGTGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1404  
QY 1615 TCCTGATCAGTGAGTGTCTGACAGTGAAGGCTCCGATATCAAAAGCAGTCCAC 1674

DB 1405 TCCTGATCAGTGAGTGTCTGACAGTGAAGGCTCCGATATCAAAAGCAGTCCAC 1464  
QY 1675 GATGATCAGTGAAGTGAATTAACCTGCAAGACCAAGATTTTAAAGGCGGCAAAAATCT 1734  
DB 1465 GATGATCAGTGAAGTGAATTAACCTGCAAGACCAAGATTTTAAAGGCGGCAAAAATCT 1524  
QY 1735 TTTCTGGAGAGTCAACAATTTGGAATGAAAAAACCAGATCAGATGTTTTATTTT 1794  
DB 1525 TTTCTGGAGAGTCAACAATTTGGAATGAAAAAACCAGATCAGATGTTTTATTTT 1584  
QY 1795 ATATTAATTAATGAGAGTGTGATCAATTAATGATGAAGGACATGACAGACCC 1854  
DB 1585 ATATTAATTAATGAGAGTGTGATCAATTAATGATGAAGGACATGACAGACCC 1644  
QY 1855 CAGCTTTGAGAGGTGTGGGGGAGAGACAGAGGAGCCCATCTGGGAAACAGCTGCAC 1914  
DB 1645 CAGCTTTGAGAGGTGTGGGGGAGAGACAGAGGAGCCCATCTGGGAAACAGCTGCAC 1704  
QY 1915 CTGGCCATGAGCTGTTTTCCCAAGATCAGTTCTGAGAGAGAGAGAGAGAGAGAGAG 1974  
DB 1705 CTGGCCATGAGCTGTTTTCCCAAGATCAGTTCTGAGAGAGAGAGAGAGAGAGAGAG 1764  
QY 1975 CCGCTGTGTCCGAGACACAGTGTGACCCGACGCCCGCTGTGATTTGGCTGGG 2034  
DB 1765 CCGCTGTGTCCGAGACACAGTGTGACCCGACGCCCGCTGTGATTTGGCTGGG 1824  
QY 2035 TCTGAGAGTGTCTGAGAGACACCGGCTCTCAACAGAGAGAGAGAGAGAGAGAGAG 2094  
DB 1825 TCTGAGAGTGTCTGAGAGACACCGGCTCTCAACAGAGAGAGAGAGAGAGAGAGAG 1884  
QY 2095 CCAAGCTGAACATGGGAAACAACCTGAAAGAGAGAGAGAGAGAGAGAGAGAGAGCT 2154  
DB 1885 CCAAGCTGAACATGGGAAACAACCTGAAAGAGAGAGAGAGAGAGAGAGAGAGAGCT 1944  
QY 2155 GCTGTGTGTGCTTCCATGACACCTCTCTGCTGAAATATTTATGCTTGAATCTGAG 2214  
DB 1945 GCTGTGTGTGCTTCCATGACACCTCTCTGCTGAAATATTTATGCTTGAATCTGAG 2004  
QY 2215 CAGATTGGGGGTTTAAACATGCTTTTATCTGAGAAACAACGGGTTGAAATTAATG 2274  
DB 2005 CAGATTGGGGGTTTAAACATGCTTTTATCTGAGAAACAACGGGTTGAAATTAATG 2064  
QY 2275 GTCTTTTTCACCACTCCAGAGAGTGTCAAGTCAATCCAGCGGCTCCGCTTGGGA 2334  
DB 2065 GTCTTTTTCACCACTCCAGAGAGTGTCAAGTCAATCCAGCGGCTCCGCTTGGGA 2124  
QY 2335 CAGGTAATGTAACTCCGATCCAGAGGCTAGCCCTGACACAGGAGGCTTCCGATCC 2394  
DB 2125 CAGGTAATGTAACTCCGATCCAGAGGCTAGCCCTGACACAGGAGGCTTCCGATCC 2184  
QY 2395 CGGTGGGAAAAGGCGCTGACACAGAGGCTTGAAGTGGCTGTGCTCCACAGCGCTG 2454  
DB 2185 CGGTGGGAAAAGGCGCTGACACAGAGGCTTGAAGTGGCTGTGCTCCACAGCGCTG 2244  
QY 2455 CACCAACCACTCCAGAGTGCAGTGTGGGCAAGGCAAGCTCAAGAGACAGAGACAGG 2514  
DB 2245 CACCAACCACTCCAGAGTGCAGTGTGGGCAAGGCAAGCTCAAGAGACAGAGACAGG 2304  
QY 2515 GCTTGGCAAGACATGACACACACCAACCAAGGCTGAGACCCAGGCTGGCGCTGG 2574  
DB 2305 GCTTGGCAAGACATGACACACCAACCAAGGCTGAGACCCAGGCTGGCGCTGG 2364  
QY 2575 TACCCAGAGGTGGGACATGCACTCCGCTCTGAGAGTCAAGGCTCTCAAGAGACA 2634  
DB 2365 TACCCAGAGGTGGGACATGCACTCCGCTCTGAGAGTCAAGGCTCTCAAGAGACA 2424  
QY 2635 CAGAGGCTGTGTCTCCGAGAGCTTCTTCAAGACCTTCTCAAGTGCACATTTGGAGT 2694  
DB 2425 CAGAGGCTGTGTCTCCGAGAGCTTCTTCAAGACCTTCTCAAGTGCACATTTGGAGT 2484  
QY 2695 CAGATGACGAGAGTGAAGTCCCTTCAAGGCTTGAAGCTTGGCTGAGTAAAGTTAC 2754



Db 2485 CAGATGACGAGGAGTACGACCCCTCCAGCGCCTGGACCTCGGCTGACGATTAAGTTAC 2544  
Qy 2755 GTGAGGCTGTCTCTCGGAGGCTTGGAAATGGCAGCCATCAATGTCTTGTGACCCCTC 2814  
Db 2545 GTGAGGCTGTCTCTCGGAGGCTTGGAAATGGCAGCCATCAATGTCTTGTGACCCCTC 2604  
Qy 2815 GGAGCAAGCCGCGACAGGTGTGTGAGACAGTGGCGGAGGAGGCCCAAGCTGGC 2874  
Db 2605 GGAGCAAGCCGCGACAGGTGTGTGAGACAGTGGCGGAGGAGGCCCAAGCTGGC 2664  
Qy 2875 CGGCTCCAGCCACCCACAGCTGTGTGAGTCAAGCCCACTCCACAGCTGTAT 2934  
Db 2665 CGGCTCCAGCCACCCACAGCTGTGTGAGTCAAGCCCTCCACAGCTGTAT 2724  
Qy 2935 CTGAGTACGGCTTAAGAACTCTCTCTCTGTGTTGAAAGCAATTGGGTTTCCAT 2994  
Db 2725 CTGAGTACGGCTTAAGAACTCTCTCTGTGTTGAAAGCAATTGGGTTTCCAT 2784  
Qy 2995 TCTGTAATTCATCTCCATTTTAAAAAAGTTTCTGTGACGGGCCCAAGGCCGAGC 3054  
Db 2785 TCTGTAATTCATCTCCATTTTAAAAAAGTTTCTGTGACGGGCCCAAGGCCGAGC 2843  
Qy 3055 CGCGGTGAGCTCGTGTGTCATGAGCCCTGGGCTTCCGTCGCTCTGCGCA 3114  
Db 2844 CGCGGTGAGCTCGTGTGTCATGAGCCCTGGGCTTCCGTCGCTCTGCGCA 2903  
Qy 3115 GGTGCTTGTGGGCAACCATCTCTGCGTTTCAATTGCACTGACGTACAGAGGCACTC 3174  
Db 2904 GGTGCTTGTGGGCAACCATCTCTGCGTTTCAATTGCACTGACGTACAGAGGCACTC 2963  
Qy 3175 ACCAATTAACCTTTCTCTGAAAGCAG 3202  
Db 2964 ACCAATTAACCTTTCTCTGAAAGCAG 2991

RESULT 5  
ADB63642  
ID ADB63642 standard; cDNA; 2454 BP.  
XX  
AC ADB63642;  
XX  
DT 04-DEC-2003 (first entry)  
XX  
DE Human cDNA encoding clone THYMU20089900.  
XX  
KW Human; ss; gene; pharmaceutical; diagnostic; gene therapy;  
KW tissue regeneration; cell regeneration; membrane protein;  
KW signal transduction-related protein; transcription-related protein;  
KW osteoporosis; neurological disease; cancer; tumour.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT 1008..1421  
FT CDS /\*tag= a  
FT /product= "Clone THYMU20089900 protein"  
XX  
EP1308459-A2.  
XX  
PD 07-MAY-2003.  
XX  
PE 28-MAR-2002; 2002EP-00007401.  
XX  
PR 05-NOV-2001; 2001JP-00379298.  
XX  
PR 25-JAN-2002; 2002US-00350978.  
XX  
PA (HELI-) HELIX RES INST.  
PA (REAS-) RES ASSOC BIOTECHNOLOGY.  
XX  
PI Isegai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S,  
PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I,  
PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Maehuo Y;  
XX

DR WPI: 2003-450961/43.  
DR P-PSDB; ADB65612.  
XX  
XX New polynucleotides and polypeptides, useful for developing a diagnostic  
PT marker or medicines for regulation of their expression and activity, or  
PT as targets of gene therapy.  
XX  
XX Claim 1; Page: 222pp; English.  
XX  
XX The invention discloses a polynucleotide comprising a sequence selected  
CC from 1970 fully defined nucleotide sequences which encode novel  
CC polypeptides. Also claimed is a polypeptide encoded by the polynucleotide  
CC or its partial peptide, an antibody binding to the polypeptide or peptide  
CC of the polynucleotide, immunologically assaying the polypeptide or  
CC peptide of the polynucleotide by contacting the polypeptide or peptide  
CC with the antibody of the encoded protein, and observing the binding  
CC between the two, a transformant carrying the polynucleotide in an  
CC expressible manner and an antisense polynucleotide. The oligonucleotide  
CC is useful as a primer for synthesizing the polynucleotide, or as a probe  
CC for detecting the polynucleotide. The polynucleotides and encoded  
CC proteins are useful as pharmaceutical agents and many disease-related  
CC genes may be included in them, for developing a diagnostic marker or  
CC medicines for regulation of their expression and activity, or as targets  
CC of gene therapy. The genes are involved in tissue and/or cell  
CC regeneration. Membrane proteins, signal transduction-related proteins,  
CC transcription-related proteins, disease-related proteins and genes  
CC neurological diseases, cancer, tumours. The cDNA may be used to regulate  
CC the activity or expression of the encoded protein to treat diseases. The  
CC sequence presented is a cDNA of the invention. Note: Some of the sequence  
CC data for this patent is not represented in the printed specification, but  
CC is based on sequence information supplied by the European Patent Office.  
XX  
SQ Sequence 2454 BP; 520 A; 707 C; 687 G; 540 T; 0 U; 0 Other;

Query Match 57.64; Score 1851.6; DB 9; Length 2454;  
Best Local Similarity 99.74; Pred. No. 0;  
Matches 1865; Conservative 0; Mismatches 4; Indels 1; Gaps 1;  
Qy 1338 GTCGCGTGAATGAGAAATGTCGGAAGATTAATGAGCAGAAAGTCCAGATAGGCG 1397  
Db 585 GTCGCGTGAATGAGAAATGTCGGAAGATTAATGAGCAGAAAGTCCAGATAGGCG 644  
Qy 1398 CTGATCATGCTCAATGCTGTGTAACGGGAAGTTGTCAATGACAAAGCAGAAAGCGAG 1457  
Db 645 CTGATCATGCTCAATGCTGTGTAACGGGAAGTTGTCAATGACAAAGCAGAAAGCGAG 704  
Qy 1458 AAGGTGAAGGTGAATGACGACGACGTCGCGCCCTGACGCTGTGAAAGGCTGGAAGCTC 1517  
Db 705 AAGGTGAAGGTGAATGACGACGACGTCGCGCCCTGACGCTGTGAAAGGCTGGAAGCTC 764  
Qy 1518 ATCCTCAGGAGGCTTCCAGAGCTGAGCTGCTGTATATGACCCGTGTGTGGGGAA 1577  
Db 765 ATCCTCAGGAGGCTTCCAGAGCTGAGCTGCTGTATATGACCCGTGTGTGGGGAA 824  
Qy 1578 GAGAGAACTGAAAGTGTCTATACAGTTCCGGGGGCTCTGCAATGAGTGTGCTG 1637  
Db 825 GAGAGAACTGAAAGTGTCTATACAGTTCCGGGGGCTCTGCAATGAGTGTGCTG 884  
Qy 1638 GACAGTGAAGGCTTCCGGAATACAAAGCAGTCCACAGATGAGTACAGATGATTAAC 1697  
Db 885 GACAGTGAAGGCTTCCGGAATACAAAGCAGTCCACAGATGAGTACAGATGATTAAC 944  
Qy 1698 GCCAAGAACCAATTTTAAAGGCGCAAAAATCTTTTCTCGGAGTCTACAAATTTG 1757  
Db 945 GCCAAGAACCAATTTTAAAGGCGCAAAAATCTTTTCTCGGAGTCTACAAATTTG 1004  
Qy 1758 GAAATGAAAAACCCAGACATCAGATGTTTATTTATTTATTTATGAAGGTGCT 1817  
Db 1005 GAAATGAAAAACCCAGACATCAGATGTTTATTTATTTATTTATGAAGGTGCT 1064  
Qy 1818 ACCAATTAATTAATGAAAGGACATGACACCCCGACTTTAGAGGTGTGCGGGGT 1877

Db 1065 ACCATTATCAATTATGTGAAGGACATGACAGACCCACAGCTTTTGAAGGTGCGGGGT 1124  
QY 1878 AGGACTGAGGAGGCCCCACCTGGGAACCACTGAGCGTGGCCCATGAGCTGTTTCCCA 1937  
Db 1125 AGGACTGAGGAGGCCCCACCTGGGAACCACTGAGCGTGGCCCATGAGCTGTTTCCCA 1184  
QY 1938 GGAATCAAGTTCTGAGAGGGAAGGCTCTGAGCTGACTCCGCTGTGTCCAGACACAGT 1997  
Db 1185 GGAATCAAGTTCTGAGAGGGAAGGCTCTGAGCTGACTCCGCTGTGTCCAGACACAGT 1244  
QY 1998 CTGACCGGACCGCGCGCTGTAGTTCTTGCGTGGGTCTGAGAGTGTCTGTGAGACAC 2057  
Db 1245 CTGACCGGACCGCGCGCTGTAGTTCTTGCGTGGGTCTGAGAGTGTCTGTGAGACAC 1304  
QY 2058 CTGCGCTCAACACAGAGCGGTGAGCCACTTCTGAGTCCAGCGTGAACATGGGAAACAC 2117  
Db 1305 CTGCGCTCAACACAGAGCGGTGAGCCACTTCTGAGTCCAGCGTGAACATGGGAAACAC 1364  
QY 2118 CTGAAAAGACGAGCGCTCCGGTCAAGGAGCGCTGTGTCTGTGCTTCCCATGACCA 2177  
Db 1365 CTGAAAAGACGAGCGCTCCGGTCAAGGAGCGCTGTGTCTGTGCTTCCCATGACCA 1424  
QY 2178 CTTCTCTCTGCTGAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2237  
Db 1425 CTTCTCTCTGCTGAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1484  
QY 2238 CTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2297  
Db 1485 CTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1544  
QY 2298 CTGCTCAAGTCAATTCACCGGCGCTCCGCTTGGGACAGGTAATTAATTCCTCCATCC 2357  
Db 1545 CTGCTCAAGTCAATTCACCGGCGCTCCGCTTGGGACAGGTAATTAATTCCTCCATCC 1604  
QY 2358 CAGGCGCTTACCTGACACAGTGTGCTTCCGCTTACCGGCGTGGGAAACCGCTGCGACC 2417  
Db 1605 CAGGCGCTTACCTGACACAGTGTGCTTCCGCTTACCGGCGTGGGAAACCGCTGCGACC 1664  
QY 2418 AGCGGCGCTTACCTGACACAGTGTGCTTCCGCTTACCGGCGTGGGAAACCGCTGCGACC 2477  
Db 1665 AGCGGCGCTTACCTGACACAGTGTGCTTCCGCTTACCGGCGTGGGAAACCGCTGCGACC 1724  
QY 2478 TGTGAGGCAAGGCGAGCTCAAGAGACAGGACAGGCGCTTGGCAAGACATCAGACACAC 2537  
Db 1725 TGTGAGGCAAGGCGAGCTCAAGAGACAGGACAGGCGCTTGGCAAGACATCAGACACAC 1784  
QY 2538 CCAACCCCAAGGCGTGAACCCGAGGCGCGCTTGTGTAACCGACAGTGTGCACTGACAC 2597  
Db 1785 CCAACCCCAAGGCGTGAACCCGAGGCGCGCTTGTGTAACCGACAGTGTGCACTGACAC 1844  
QY 2598 TCCCGGCTCTGAGAGTCCAGCGTCTTCAAGAAACAGAGGCGTGTGCTCCGAGCGCT 2657  
Db 1845 TCCCGGCTCTGAGAGTCCAGCGTCTTCAAGAAACAGAGGCGTGTGCTCCGAGCGCT 1904  
QY 2658 TCCCTTCAAGCCCTTCTTCAAGTCCGCTTGGGATGCAAGATGACAGCGAGCTAGACAC 2717  
Db 1905 TCCCTTCAAGCCCTTCTTCAAGTCCGCTTGGGATGCAAGATGACAGCGAGCTAGACAC 1964  
QY 2718 CCGTCAAGCGGCTGAGACTTGGCTGCAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 2777  
Db 1965 CCGTCAAGCGGCTGAGACTTGGCTGCAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 2024  
QY 2778 GGAAGTGAAGCCATCAAGTGTCTTGTGTAACCCCTGAGAGAGCGCCAGAGGTGGT 2837  
Db 2025 GGAAGTGAAGCCATCAAGTGTCTTGTGTAACCCCTGAGAGAGCGCCAGAGGTGGT 2084  
QY 2838 GAGTGAAGCAGCTGCGCGGAGGAGCCCAAGCTGCGCGGCTTCCAGCCACCAAGCT 2897  
Db 2085 GAGTGAAGCAGCTGCGCGGAGGAGCCCAAGCTGCGCGGCTTCCAGCCACCAAGCT 2144  
QY 2898 GTTGTGAAGTCAAGGCGCAACTCCAGGCACTGTGTATGTGAAGTCAAGGCGCAACTCC 2957  
Db 2145 GTTGTGAAGTCAAGGCGCAACTCCAGGCACTGTGTATGTGAAGTCAAGGCGCAACTCC 2204

QY 2958 TTCTCTGTTTTGAAAAGCAGTTGCGGTTGTCTCAATTTCTGTAACATTCATCCATTT 3017  
Db 2205 TTCTCTGCGTTTGAAGAACAGTTGCGGTTGTCTCAATTTCTGTAACATTCATCCATTT 2264  
QY 3018 TTAAGAAAGTTTCTTGAACGCGCCCAAGCGCGCGGTGAGCGTGTGTGATG 3077  
Db 2265 TT-TAAGAAAGTTTCTTGAACGCGCCCAAGCGCGCGGTGAGCGTGTGTGATG 2323  
QY 3078 AGCTGGGCGCGCGGTTTCCGTTGCGCTCTGCGCGAGTGTCTTGTGGACATCCCTC 3137  
Db 2324 AGCTGGGCGCGCGGTTTCCGTTGCGCTCTGCGCGAGTGTCTTGTGGACATCCCTC 2383  
QY 3138 TGGTTTCAATTTGACAGTGTGACAGAGGCACTCACCAATTAACCTTTCTGAA 3197  
Db 2384 TGGTTTCAATTTGACAGTGTGACAGAGGCACTCACCAATTAACCTTTCTGAA 2443  
QY 3198 GCAGAAAAA 3207  
Db 2444 GCAGAAAAA 2453

RESULT 6  
AAK91227  
ID AAK91227 standard; DNA; 19701 BP.  
XX  
AC AAK91227;  
XX  
XX 05-NOV-2001 (first entry)  
XX  
DE Human digestive system antigen genomic sequence SEQ ID NO: 4803.  
XX  
XX  
KM Human; digestive system antigen; gene therapy; cancer; appendicitis;  
KM ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;  
KM digestive system disorder; Meckel's diverticulum; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200155314-A2.  
XX  
PD 02-AUG-2001.  
XX  
XX  
PF 17-JAN-2001; 2001WO-US001324.  
XX  
PR 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-019076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.



QY	18449	TCACCCAGCMTTGTGAGGGTGTCTGGGGGTGAGATCTAGAGGAGGCCCACTGGGAACCAAGAC	1908
Db	18348	ACACCCCAAGCTTTTGTAGAGGGTGTCTGGGGGTGAGATCTAGAGGAGGCCCACTGGGAACCAAGAC	18407
QY	1909	TGCAGCTCGGGCCCAATGGGTGTGTTTTCCCAAGGATCAATCTCTGGAGGGAAAGGGCTCTGGGC	1968
Db	18408	TGCAGCTCGGGCCCAATGGGTGTGTTTTCCCAAGGATCAATCTCTGGAGGGAAAGGGCTCTGGGC	18467
QY	1969	CTGACTCCGCTGTGTCTCCGAGCACAGTGTGACCCGAGGCCCGCCCTGTAGTCTTTG	2028
Db	18468	CTGACTCCGCTGTGTCTCCGAGCACAGTGTGACCCGAGGCCCGCCCTGTAGTCTTTG	18527
QY	2029	GCTGGGTGTGGAGGTGTCTGTGGAGCACCTGTGCCCTTACCAACAGAGGCTGAGCCATTC	2088
Db	18528	GCTGGGTGTGGAGGTGTCTGTGGAGCACCTGTGCCCTTACCAACAGAGGCTGAGCCATTC	18587
QY	2089	TGCAGTCCACGCTTGAACATGGGGAACAACTGAAACAGAGGAGGAGGCTCCCGGTCAAGGA	2148
Db	18588	TGCAGTCCACGCTTGAACATGGGGAACAACTGAAACAGAGGAGGAGGCTCCCGGTCAAGGA	18647
QY	2149	GCTCTGTGTGTGTGGCTTCCCATGACCACTCTCTGTGTGAATATTAATCTGCTTGAAT	2208
Db	18648	GCTCTGTGTGTGTGGCTTCCCATGACCACTCTCTGTGTGAATATTAATCTGCTTGAAT	18707
QY	2209	CTGAGAGCAGATTTGGGGGTTTATTAATACTGCTTTTATCTGAGAAACAAAGGGTTTGGAA	2268
Db	18708	CTGAGAGCAGATTTGGGGGTTTATTAATACTGCTTTTATCTGAGAAACAAAGGGTTTGGAA	18767
QY	2269	TTAGTCGTCTTTTATTTCCCACTCCCAAGAGTGTCAAGTCAATTCACCGGACCCCTCGGC	2328
Db	18768	TTAGTCGTCTTTTATTTCCCACTCCCAAGAGTGTCAAGTCAATTCACCGGACCCCTCGGC	18827
QY	2329	TTGGAGCAGGGGTAGTGTAACTCCCGATCCAGGGCCTTAGCCCTGACACAGTGGCTTCCC	2388
Db	18828	TTGGAGCAGGGGTAGTGTAACTCCCGATCCAGGGCCTTAGCCCTGACACAGTGGCTTCCC	18887
QY	2389	GTATTCCTCGGTGGGAAAAAGCCCTGTCCACACAGCGGGCTTGGAGCTGTGTCTCTCCAC	2448
Db	18888	GTATTCCTCGGTGGGAAAAAGCCCTGTCCACACAGCGGGCTTGGAGCTGTGTCTCTCCAC	18947
QY	2449	CGCCTGACACCAACCCACCTCCAGAGTGTGAGGCTGGGCAAGGGAGGCTCAAGAGGACAGGA	2508
Db	18948	CGCCTGACACCAACCCACCTCCAGAGTGTGAGGCTGGGCAAGGGAGGCTCAAGAGGACAGGA	19007
QY	2509	CCAGGCGCTTGGCAAGACATCAGACACACCACCCAAAGGCGTGGACCCCAAGCCCGGC	2568
Db	19008	CCAGGCGCTTGGCAAGACATCAGACACACCACCCAAAGGCGTGGACCCCAAGCCCGGC	19067
QY	2569	CCGTGTGTATCCCAAGAGGTGGCATCTGACAGCTCCCGCTCTGTGCAGGCTCAGCTCTTACA	2628
Db	19068	CCGTGTGTATCCCAAGAGGTGGCATCTGACAGCTCCCGCTCTGTGCAGGCTCAGCTCTTACA	19127
QY	2629	GGAACACCAAGGGCTGTGTCTCCGAGAGCCTTCTTACAGACCTTCTCTCAAGTGGCCACAT	2688
Db	19128	GGAACACCAAGGGCTGTGTCTCCGAGAGCCTTCTTACAGACCTTCTCTCAAGTGGCCACAT	19187
QY	2689	GGAGTGCAGAAATGACAGCGAGCTTAGGACCCCTTCCACAGGCTTGGACCTCGCTGCAGTAA	2748
Db	19188	GGAGTGCAGAAATGACAGCGAGCTTAGGACCCCTTCCACAGGCTTGGACCTCGCTGCAGTAA	19247
QY	2749	AGTTATCTGAGAGGCTGTGTCTCTCGGGGCTTGGAAAGTGCAGCCATCAATGTGTCTTGTGA	2808
Db	19248	AGTTATCTGAGAGGCTGTGTCTCTCGGGGCTTGGAAAGTGCAGCCATCAATGTGTCTTGTGA	19307
QY	2809	CCCTTCGGAGCAAGCGCGCACAGGTGTGTGTGAGCAGCTGGCGCGGGGGGGCCCAAG	2868
Db	19308	CCCTTCGGAGCAAGCGCGCACAGGTGTGTGTGAGCAGCTGGCGCGGGGGGGCCCAAG	19367
QY	2869	CTGGCGCGGCTCTCAGGCCACCAACCACTGTGTGCTGAAGTCAAGGCAACCTCCCAAGCAC	2928
Db	19368	CTGGCGCGGCTCTCAGGCCACCAACCACTGTGTGCTGAAGTCAAGGCAACCTCCCAAGCAC	19427
QY	2929	TGATATCTAGTAAAGCGCTAGAAACCTCTTCTGTGTGTTTGAAGAGAGTTCCGGTTG	2988

Db	19428	TGATATCTGAGTAAAGGCTAAGAACTCTTCTCTGCTTTTGAAAGAGATTCCGGTTG	19487
Oy	2989	TCCAATTCGTGAACATTCATCTCCATTTTTTAAAAAGGTTTCTGACGGCCCAACGGC	3048
Db	19488	TCCAATTCGTGAACATTCATCTCAATTTTTT-AAAAAGGTTTCTGACGGCCCAACGGC	19546
Oy	3049	CCGAGCCGCGGTGAGCGCTGCTGTTCAGTAGAGCTGGCCCCCGGACTTCCCGTCCGCTCT	3108
Db	19547	CCGAGCCGCGGTGAGCGCTGCTGTTCAGTAGAGCTGGCCCCCGGACTTCCCGTCCGCTCT	19606
Oy	3109	GCCCGAGTGTCTCTGGGACCCCATCTCTCGCTTCATTTGACAGTCGACTGTACAGAG	3168
Db	19607	GCCCGAGTGTCTCTGGGACCCCATCTCTCGCTTCATTTGACAGTCGACTGTACAGAG	19666
Oy	3169	GCACTCACCACCAATTAACCTTTCTCTGAAAGCAGA	3202
Db	19667	GCACTCACCACCAATTAACCTTTCTCTGAAAGCAGA	19700
RESULT 7			
AB054721			
ID	AB054721	standard; cDNA; 1454 BP.	
XX	AB054721;		
AC	22-AUG-2002	(first entry)	
XX			
DE	Human ovarian antigen HPAME01 cDNA, SEQ ID NO:601.		
XX			
KW	Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;		
KW	ovarian cancer; breast cancer; tumour; reproductive system disorder;		
KW	infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;		
KW	PCOS; ovarian cyst; dysmenorrhoea; endocrine disorder; infection;		
KW	inflammatory condition; immune disorder; blood disorder;		
KW	cardiovascular disorder; respiratory disorder; neurological disorder;		
KW	gastrointestinal disorder; urinary system disorder; drug screening;		
KW	gene therapy; chromosome mapping; forensic analysis;		
KW	antibody preparation; cytostatic; immunomodulatory; neuroprotective;		
KW	antiinflammatory; gynaecological; reproductive; gene; ss.		
XX			
OS	Homo sapiens.		
XX			
PN	WO200200677-A1.		
XX			
PD	03-JAN-2002.		
XX			
PF	07-JUN-2001; 2001MO-US018569.		
XX			
PR	07-JUN-2000; 2000US-0209467P.		
XX			
PA	(HUMA-) HUMAN GENOME SCI INC.		
XX			
PI	Birse CE, Rosen CA;		
XX			
DR	WPI; 2002-147878/19.		
XX	P-PSDB; ABP41644.		
PT	Isolated nucleic acid molecules encoding novel ovarian polypeptides,		
PT	useful in the prevention, treatment and diagnosis of cancer (e.g. ovarian		
PT	cancer), immune disorders, cardiovascular disorders and neurological		
XX	diseases.		
PS	Claim 1; SEQ ID NO 601; 2922pp; English.		
XX			
CC	The invention relates to 2175 novel human ovarian antigens (ABP41054-		
CC	ABP43228) and to cDNAs encoding them (AB054131-AB056305), and also		
CC	encompasses polypeptides 90% identical and polynucleotides 95% identical		
CC	to the sequences of the invention. The invention additionally relates to		
CC	recombinant vectors and host cells comprising human ovarian antigen		
CC	polynucleotides, antibodies against human ovarian antigens, and the use		
CC	of ovarian antigen polynucleotides and polypeptides in diagnosing,		
CC	treating, prognosing or preventing various ovary and/or breast-related		



PD		21-SEP--2000.	
PF		08-MAR--2000; 2000WO-US005988.	
XX		12-MAR-1999; 99US-0124270P.	
PR		(HUMA-) HUMAN GENOME SCI INC.	
XX		(ROSE/) ROSEN C A.	
PA		Rosen CA, Ruben SM;	
PI		WPI; 2000-587513/55.	
DR		P-PsDB; AAB56849.	
XX			
PT		Prostate cancer associated gene sequences, referred to as prostate cancer	
PT		antigens, useful for treatment, prevention, and diagnosis of disorders	
XX		such as prostate cancer.	
PS		Claim 1; Page 957-958; 238bp; English.	
CC		AAP1556 to AAP1505 encode the human prostate cancer associated	
CC		proteins, called prostate cancer antigens, given in AAB563 to AAB5702.	
CC		The prostate cancer antigens can have neuroprotective, cytostatic,	
CC		cardioactive, immunomodulatory, muscular, vulnary, gastrointestinal,	
CC		nephrotropic, antileffective, gynaecological and antibacterial activities,	
CC		and can be used in gene therapy. The prostate cancer antigen	
CC		polynucleotides may be used for detection of prostate cancer, chromosome	
CC		identification, as chromosome markers, and for numerous other diagnostic	
CC		or research purposes. The prostate cancer antigens may be used to treat	
CC		disorders such as neural, immune, muscular, reproductive,	
CC		gastrointestinal, pulmonary, cardiovascular, renal, and proliferative	
CC		disorders, wounds, and infectious diseases. AAP1506 to AAP1514 to	
CC		AAB5703 represent sequences used in the exemplification of the present	
CC		invention	
SQ		Sequence 1451 BP; 308 A; 449 C; 386 G; 304 T; 0 U; 4 Other;	
	Query Match	44.0%; Score 1413.6; DB 3; Length 1451;	
	Best Local Similarity	99.5%; Pred. No. 0;	
	Matches 1424; Conservative	3; Mismatches 3; Indels 1; Gaps 1;	
OY	1783	TGTTTTATTATTAATTTAATTAATAGAAAGTGTAACATTATCATTTATGTGAAGGCAC	1842
Db	1	TGTTTTATTATTAATTTAATTAATAGAAAGTGGTACCAATTATCATTTATGTGAAGGCAC	60
OY	1843	ATGCAGAACCACCCAGCTTTTAGAGGGTGTCTGGGGGGTAGGACTGAGGAGACCCCACCTGGGAA	1902
Db	61	ATGCAGAACCACCCAGCTTTTAGAGGGTGTCTGGGGGGTAGGACTGAGGAGACCCCACCTGGGAA	120
OY	1903	CCAGACTCAGCTGGGCCCATGTGCTTTTCCCAAAGATTCAGTTCTTGAAGGGAAGGCT	1962
Db	121	CCAGACTCAGCTGGGCCCATGTGCTTTTCCCAAAGATTCAGTTCTTGAAGGGAAGGCT	180
OY	1963	CTGGCCCTGACTCGCGCTGTGTCGCCGAGACAACGTGCTGACCGGACCCGCGCCTGTAG	2022
Db	181	CTGGCCCTGACTCGCGCTGTGTCGCCGAGACAACGTGCTGACCGGACCCGCGCCTGTAG	240
OY	2023	TTCTTGGCTGGGTCTGGAAGGTGTCTGTGGAAGAACCTTCCCTCACAGAGGGTGAAC	2082
Db	241	TTCTTGGCTGGGTCTGGAAGGTGTCTGTGGAAGAACCTTCCCTCACAGAGGGTGAAC	300
OY	2083	CACTTCTSCAGTCCAAGCTGAAACATGGAGAAAACAACCTTGAAGGAGGAGGCTTCCGGT	2142
Db	301	CACTTCTSCAGTCCAAGCTGAAACATGGAGAAAACAACCTTGAAGGAGGAGGCTTCCGGT	360
OY	2143	CAGGAGGCTCTGCTGTGCTGTGCTTCCATGACCACTTCTTCTGCTGAATAATTAATCTGC	2202
Db	361	CAGGAGGCTCTGCTGTGCTGTGCTTCCATGACCACTTCTTCTGCTGAATAATTAATCTGC	420
OY	2203	TTGAATCTGAGACAAATTGGGGTTTATTAATAATGCTTTTATCTGAGAAACAACGGGTT	2262
Db	421	TTGAATCTGAGACAAATTGGGGTTTATTAATAATGCTTTTATCTGAGAAACAACGGGTT	480

QY	2263	TGGAATTAAGTCGCTTTTTCACCATCCCGAGAGCTGCTCAATCATATCCAGCGGCC	2332
Db	461	TGGAATTAAGTCGCTTTTTCACCATCCCGAGAGCTGCTCAATCATATCCAGCGGCC	540
QY	2333	CTCGGCTTGGACAAGGGTAGTGTAACTCCGATCCAGAGGCTTACCCCTGACACAGGTGG	2382
Db	541	CTCGGCTTGGACAAGGGTAGTGTAACTCCGATCCAGAGGCTTACCCCTGACACAGGTGG	600
QY	2383	CTTCCCGTATCCGAGTGGAAAAAGCCCTGACACAGCGGCTTGAAGCTGGCTGTGTCC	2442
Db	601	CTTCCCGTATCCGAGTGGAAAAAGCCCTGACACAGCGGCTTGAAGCTGGCTGTGTCC	660
QY	2443	CTCCACGCGCTGACACACCCACTCTCCAGAGTGTGTCTGGGGCAAGGGCAGCTTAAGAG	2502
Db	661	CTCCACGCGCTGACACACCCACTCTCCAGAGTGTGTCTGGGGCAAGGGCAGCTTAAGAG	720
QY	2503	ACAGGACCAAGGCGCTTGGCAAGACATAGACACACCCAAAGGCGTGTGACCCAGG	2562
Db	721	ACAGGACCAAGGCGCTTGGCAAGACATAGACACACCCAAAGGCGTGTGACCCAGG	780
QY	2563	CCCGGCGCGTGTGTACCCAGCAGGTGTGACATGTCCGCTCTGTGACAGTTCAGGCTC	2622
Db	781	CCCGGCGCGTGTGTACCCAGCAGGTGTGACATGTCCGCTCTGTGACAGTTCAGGCTC	840
QY	2623	CTCACAGGAACAACAGAGGCTGTGTCTCCGAGAGCTTCTCTTAAGACCTTCTCTCAAGTGC	2682
Db	841	CTCACAGGAACAACAGAGGCTGTGTCTCCGAGAGCTTCTCTTAAGACCTTCTCTCAAGTGC	900
QY	2683	CCACTTGGGATGTAGAAATGTACAGCGGAGCTAGGACCCCTCCAGCGGCTGTGACCTCGGCTG	2742
Db	901	CCACTTGGGATGTAGAAATGTACAGCGGAGCTAGGACCCCTCCAGCGGCTGTGACCTCGGCTG	960
QY	2743	CAGTAAAGTTACGTGAGGCGTGTCTCTCGGGGCTGTGAAGTGTGAGGCATTCAGTTGTCT	2802
Db	961	CAGTAAAGTTACGTGAGGCGTGTCTCTCGGGGCTGTGAAGTGTGAGGCATTCAGTTGTCT	1020
QY	2803	TGCTGACCCCTCGAGACAGCGCGGCACAGGTGTGTGTGAGACAGCTGTGCGCGGGGCGC	2862
Db	1021	TGCTGACCCCTCGAGACAGCGCGGCACAGGTGTGTGTGAGACAGCTGTGCGCGGGGCGC	1080
QY	2863	CCCAAGCTGTGCGCGGCGCTCCAGGCCACCCACAGCTGTGTGTGAGTCAAGGCACACTCCG	2922
Db	1081	CCCAAGCTGTGCGCGGCGCTCCAGGCCACCCACAGCTGTGTGTGAGTCAAGGCACACTCCG	1140
QY	2923	CAGCACTGTGTATCTGAGTAACGGCTAAGAACCTTCTCTGTGTTTGAAGAACAGTTC	2982
Db	1141	CAGCACTGTGTATCTGAGTAACGGCTAAGAACCTTCTCTGTGTTTGAAGAACAGTTC	1200
QY	2983	GGGTTGTCCAAATCTGTACACTTCATCTCCATTTTAAAAAAGTTTCTCTGACGCGCC	3042
Db	1201	GGGTTGTGTCCAAATCTGTACACTTCATCTCCATTTTAAAAAAGTTTCTCTGACGCGCC	1259
QY	3043	CACGCGCCGAGCGCGGAGCGGTGTGTGTGCAAGACCTGTGGGCGCGGCGTTCCGCGTC	3102
Db	1260	CACGCGCCGAGCGCGGAGCGGTGTGTGTGCAAGACCTGTGGGCGCGGCGTTCCGCGTC	1319
QY	3103	GCTCTGTGCGGAGGTGTCTTCTGTGGGACCCCATCTCTCGTTTCATTTGACAGTCAGCTGA	3162
Db	1320	GCTCTGTGCGGAGGTGTCTTCTGTGGGACCCCATCTCTCGTTTCATTTGACAGTCAGCTGA	1379
QY	3163	CAGAAAGCACTCACCAATTAACCTTTCCTGAAGAGCAAAAAAAGAAAAA 3213	
Db	1380	CAGAAAGCACTCACCAATTAACCTTTCCTGAAGAGCAAAAAAAGAAAAA 1430	
RESULT 9			
ABS64607			
ID	ABS64607	standard; cDNA; 1014 BP.	
XX	AC	ABS64607;	
XX	DT	15-NOV-2002 (first entry)	

Human cDNA probe selectively hybridizing to foetal cell mRNA #22.

Human: ss: foetal liver myeloid cell; probe: erythroblast; foetal abnormality; maternal blood; differential display; chromosome1 triplet expansion disorder; aneuploidy; 21; Klinefelter syndrome; spina bifida; sickle cell anaemia; thalasassaemia; Marfan syndrome; Duchenne muscular dystrophy; cystic fibrosis; Fragile X-syndrome; Friedrich's ataxia; myotonic dystrophy; Huntington's disease; sex determination.

Homo sapiens.

MO20255985-A2.

18-JUL-2002.

01-NOV-2001; 2001WO-US045340.

15-NOV-2000; 2000US-0248882P.

(HOPF ) ROCHE DIAGNOSTICS CORP.

Schneier PA, Xu H, Foltz L, Wu X, Sha Y, Nagy A, Mahoney WC; WPI; 2002-619108/66.

P-Psdb; ABG79327.

Detecting fetal cells in maternal blood sample, useful for diagnosing an abnormality in a fetal cell, involves using specific nucleic acid probes that hybridize to fetal cell associated RNAs.

Claim 109; Page 168-169; 215pp; English.

The invention relates to detecting foetal cells in maternal blood sample, involving contacting sample with a first probe or performing expression analysis on RNA/cDNA obtained from foetal liver myeloid cells relative to that obtained from mature cells to identify RNA/cDNA species useful as probe, contacting sample with the probe and identifying if the sample comprises a cell that comprises mRNA that hybridises to the probe. The probes are identified by differential display analysis using mature liver cells and foetal liver myeloid cells of less than 22 weeks of gestation. Also included are: (1) The probe sequences (appearing as AB664586-AB664618), an isolated nucleic acid molecule having a sequence which is at least 90% identical to the probe or its complement, or having a nucleotide sequence identical to at least 15 consecutive nucleotide residues of the probe, the encoded proteins from the probe or encoding a naturally occurring allelic variant or fragment; (2) a non-mammalian host cell containing the probe and (3) an antibody which selectively binds to the probe encoded protein. The method is useful for detecting a foetal cell (such as erythroblast or trophoblast) in a maternal blood sample, which is useful for diagnosing a chromosomal abnormality, single gene disorder or nucleotide triplet expansion in the gene, in a foetal cell. For example an aneuploidy (trisomy 13, trisomy 21, or Klinefelter syndrome), spina bifida, sickle cell anaemia, a thalasassaemia, Marfan syndrome, Duchenne muscular dystrophy, cystic fibrosis, Fragile X-syndrome, Friedrich's ataxia, myotonic dystrophy or Huntington's disease. The probes are also useful for determination of the sex of a fetus and for detecting the presence of multiple foetuses at early stages of pregnancy. The present sequence is a foetal cell specific probe of the invention

Sequence 1014 BP; 275 A; 242 C; 284 G; 213 T; 0 U; 0 Other;

Query Match 31.3%; Score 1006.6; DB 6; Length 1014; Best Local Similarity 99.6%; Pred. No. 5.8e-280; Matches 1009; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

11 GAAATGCGCAGCGCTTGTAGCGAGAGACTGACAAATGAAGCTATTACTGCTTCT 70  
|||||  
2 GAAATGCGCAGCGCTTGTAGCGAGAGACTGACAAATGAAGCTATTACTGCTTCT 61  
|||||

71 GAACTGCGCAGCGCTTGTAGAGAGCTGAAGAGCTCTACCGAGGCTCTGTAT 130  
|||||

62 GAACTGCGCAGCGCTTGTAGAGAGCTGAAGAGCTCTACCGAGGCTCTGTAT 121  
|||||

131 GCTTACCATTCAGACAGACAGAGCCAGAGCTCAAGTCACAGGCGGAGACACTGTT 190  
|||||

122 GCTTACCATTCAGACAGACAGAGCCAGAGCTCAAGTCACAGGCGGAGACACTGTT 181  
|||||

191 TAACCTGTTACAGAGCTTATGAAGCTTATGAGACCCCAAAACAGGGCCATCTATGA 250  
|||||

182 TAACCTGTTACAGAGCTTATGAAGCTTATGAGACCCCAAAACAGGGCCATCTATGA 241  
|||||

251 TATATATGGAGAGAGAGACTGGAATATGAAGATGAGAGGCTTGTGAAAAGAGAGAAC 310  
|||||

242 TATATATGGAGAGAGAGACTGGAATATGAAGATGAGAGGCTTGTGAAAAGAGAGAAC 301  
|||||

311 CCTGCTGAATTTGAGAGAGATTGAGCGGCTTCAGAGAGAGAGAGAGAGAGATT 370  
|||||

302 CCTGCTGAATTTGAGAGAGATTGAGCGGCTTCAGAGAGAGAGAGAGAGAGATT 361  
|||||

371 GCAGCAGCGAAGCAATCCCAAGGAGATCAGGCTTGAAGTGAATCCAGCCCTTTT 430  
|||||

362 GCAGCAGCGAAGCAATCCCAAGGAGATCAGGCTTGAAGTGAATCCAGCCCTTTT 421  
|||||

431 TGATCGCTATGATGAGAGATGAGAGATGATGTCGCGAGTAGCTTTCCGAGATTGAAT 490  
|||||

422 TGATCGCTATGATGAGAGATGAGAGATGATGTCGCGAGTAGCTTTCCGAGATTGAAT 481  
|||||

491 TAATATAATGACATATATCCAGTTCATTTAGAGACCTTTGACAGAGACAGACCAT 550  
|||||

482 TAATATAATGACATATATCCAGTTCATTTAGAGACCTTTGACAGAGACAGACCAT 541  
|||||

551 CCTCTGGAAGGCTCTCAACCCAGAGATGAGAGAGGTTCCATTAATCTTTGGCT 610  
|||||

542 CCTCTGGAAGGCTCTCAACCCAGAGATGAGAGAGGTTCCATTAATCTTTGGCT 601  
|||||

611 CAGACGAGTAATCTTCGCAAAAGGATGAGAGAGATTGGAATTTGAGACTGAGACCTTACA 670  
|||||

602 CAGACGAGTAATCTTCGTAAGAGATGAGAGAGATTGGAATTTGAGACTGAGACCTTACA 661  
|||||

671 GGGGCTTTGTTGGTCTCAAGCTGTCCTGATCTCAACCAAGATGCTTTGTGACAC 730  
|||||

662 GGGGCTTTGTTGGTCTCAAGCTGTCCTGATCTCAACCAAGATGCTTTGTGACAC 721  
|||||

731 AAATGTCCTGAGATTTTCAATCCGTGAATCCGACCGGCTGACACTGCTTACG 790  
|||||

722 AAATGTCCTGAGATTTTCAATCCGTGAATCCGACCGGCTGACACTGCTTACG 781  
|||||

791 TCGGAACCTAGACAAAGAACCGTGGCTACTGCACTGAGATGGGTTATCCAGTCAGC 850  
|||||

782 TCGGAACCTAGACAAAGAACCGTGGCTACTGCACTGAGATGGGTTATCCAGTCAGC 841  
|||||

851 CATGAACATGACATGCTGCGAGACACTAAACCGACACTTCACTGAGCCCTGACGT 910  
|||||

842 CATGAACATGACATGCTGCGAGACACTAAACCGACACTTCACTGAGCCCTGACGT 901  
|||||

911 GGGAAATCCCTCACTCTTTTGACATGATCACTATCAGACAAATTTCCAATATGACATCA 970  
|||||

902 GGGAAATCCCTCACTCTTTTGACATGATCACTATCAGACAAATTTCCAATATGACATCA 961  
|||||

971 GACTCGTGAAGAGATCCCTCAAGAGAGCTTTTGGAGCGGTGTGAGT 1023  
|||||

962 GACTCGTGAAGAGATCCCTCAAGAGAGCTTTTGGAGCGGTGTGAGT 1014  
|||||

RESULT 10  
AA160714/c  
ID AA160714 standard; cDNA, 1593 BP.  
XX  
AC AA160714;  
XX  
XX 22-Oct-2001 (first entry)  
XX Human polynucleotide SEQ ID NO 4703.



XX Human; nootropic; immunosuppressant; cytoskeletal; gene therapy; cancer;  
KM peripheral nervous system; neuropathy; central nervous system; CNS;  
KM Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
KM amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
KM chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
KM leukaemia; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200153312-A1.  
XX  
PD 26-JUL-2001.  
XX  
PF 26-DEC-2000; 2000MO-US0344263.  
XX  
PR 23-DEC-1999; 99US-00471275.  
PR 21-JAN-2000; 2000US-00488725.  
PR 25-APR-2000; 2000US-00552317.  
PR 20-JUN-2000; 2000US-00598042.  
PR 19-JUL-2000; 2000US-00620312.  
PR 03-AUG-2000; 2000US-00653450.  
PR 14-SEP-2000; 2000US-00662191.  
PR 19-OCT-2000; 2000US-00693036.  
PR 29-NOV-2000; 2000US-00727344.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D,  
PI Wang J, Wang Z, Wehman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao Q;  
PI Zhou P, Goodrich R, Drmanac RT;  
XX  
XX WPI; 2001-442253/47.  
DR P-PSDB; AAM41558.  
XX  
PT Novel nucleic acids and polypeptides, useful for treating disorders such  
PT as central nervous system injuries.  
XX  
PS Claim 1; SEQ ID NO 4703; 10078bp; English.  
XX  
XX The invention relates to human nucleic acids (AA157798-AA161369) and the  
CC encoded polypeptides (AAM38642-AAM42213) with nootropic,  
CC immunosuppressant and cytoskeletal activity. The polynucleotides are useful  
CC in gene therapy. A composition containing a polypeptide or polynucleotide  
CC of the invention may be used to treat diseases of the peripheral nervous  
CC system, such as peripheral nervous injuries, peripheral neuropathy and  
CC localised neuropathies and central nervous system diseases, such as  
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
CC utilisation of the activities such as: Immune system suppression,  
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening  
CC assays for receptor activity, arthritis and inflammation, leukaemias and  
CC C.N.S disorders. Note: The sequence data for this patent did not form  
CC part of the printed specification  
XX  
SQ Sequence 1593 BP; 336 A; 445 C; 435 G; 375 T; 0 U; 2 Other;

Query Match 25.9%; Score 832.2; DB 4; Length 1593;  
Best Local Similarity 99.1%; Pred. No. 2.2e-229;  
Matches 83; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 GAAAGTTGCGAAGATGCGACGAGCTTGTGACGAGAGAGCTGCAATGAAGACTAAT 60  
DB 1455 GAAAGTTGCGAAGATGCGACGAGCTTGTGACGAGAGAGCTGCAATGAAGACTAAT 1396  
QY 61 ACTGTTGCTGAAAGTGCAGGAGGCTCTTTCTGAAGCTGAAAGCTGCTTACCGGA 120  
DB 1395 ACTGTTGCTGAAAGTGCAGGAGGCTCTTTCTGAAGCTGAAAGCTGCTTACCGGA 1336  
QY 121 GGCTCTGATGCTCTACCATCCAGACAGACAGACAGACCCAGAGCTCAAGTCAAGGCGG 180  
DB 1335 GGCTCTGATGCTCTACCATCCAGACAGACAGACAGACCCAGAGCTCAAGTCAAGGCGG 1276

QY 181 AACGACTGTTTAACCTTGTTCACAGGCTTATGAAGTCTTAGTGAACCCCAACGAGG 240  
DB 1275 AACGACTGTTTAACCTTGTTCACAGGCTTATGAAGTCTTAGTGAACCCCAACGAGG 1216  
QY 241 CCATCTATATATATATATGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300  
DB 1215 CCATCTATATATATATATGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1156  
QY 301 GGAGAGAAACCCCTGCTGAATTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360  
DB 1155 GGAGAGAAACCCCTGCTGAATTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1096  
QY 361 AGAGAGATTTGAG 420  
DB 1095 AGAGAGATTTGAG 1036  
QY 421 CCGACTTTTGTATGCTATATGATGAGAGATATGAAGATGTGCCGAGTACTTCCGC 480  
DB 1035 CCGACTTTTGTATGCTATATGATGAGAGATATGAAGATGTGCCGAGTACTTCCGC 976  
QY 481 AGATTGAATTAATTAATGAACATATCCAGTCCATTGAGGACCCCTTGACAGAGAGAG 540  
DB 975 AGATTGAATTAATTAATGAACATATCCAGTCCATTGAGGACCCCTTGACAGAGAGAG 916  
QY 541 ACACAGCATCTCTCTGGAAGCTCTCAACCCAGATGAATGAATGAAGAGAGAGAGAGAGAG 600  
DB 915 ACACAGCATCTCTCTGGAAGCTCTCAACCCAGATGAATGAATGAAGAGAGAGAGAGAG 856  
QY 601 ACTTTGCGCTCAGACGAGATTAATCTTGGCAAGAGAGATGGGAGAGAGAGAGAGAGAG 660  
DB 855 ACTTTGCGCTCAGACGAGATTAATCTTGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 796  
QY 661 GAGACCTAAGGAGGCTTGTGGCTCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 720  
DB 795 GAGACCTAAGGAGGCTTGTGGCTCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 736  
QY 721 TTGTGACAAACAATGCTGCTGCAAGTTTTCATCCCGTGAATCCGACCCGAGCTGACCA 780  
DB 735 TTGTGACAAACAATGCTGCTGCAAGTTTTCATCCCGTGAATCCGACCCGAGCTGACCA 676  
QY 781 CTGTCTTACCTGGAACCTTGAACAGACAGACAGACAGACAGACAGACAGACAGACAG 840  
DB 675 CTGTCTTACCTGGAACCTTGAACAGACAGACAGACAGACAGACAGACAGACAGACAG 616  
QY 841 TCCAG 845  
DB 615 CCCCC 611

RESULT 11  
AA158928  
ID AA158928 standard; cDNA; 1767 BP.  
XX  
AC AA158928;  
XX  
XX  
DT 22-OCT-2001 (first entry)  
XX  
XX Human polynucleotide SEQ ID NO 1131.  
XX  
XX Human; nootropic; immunosuppressant; cytoskeletal; gene therapy; cancer;  
KM peripheral nervous system; neuropathy; central nervous system; CNS;  
KM Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
KM amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
KM chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
KM leukaemia; ss.  
XX  
XX Homo sapiens.  
XX  
OS  
XX  
PN WO200153312-A1.  
XX  
PD 26-JUL-2001.  
XX



PF 26-DEC-2000; 2000WO-US034263.  
 XX 23-DEC-1999; 99US-00471275.  
 PR 21-JAN-2000; 2000US-0048725.  
 PR 25-APR-2000; 2000US-00552317.  
 PR 20-JUN-2000; 2000US-00598042.  
 PR 19-JUL-2000; 2000US-00620312.  
 PR 03-AUG-2000; 2000US-00653450.  
 PR 14-SEP-2000; 2000US-00662191.  
 PR 19-OCT-2000; 2000US-00693036.  
 PR 29-NOV-2000; 2000US-00727344.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Tang YT, Liu C, Neundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D,  
 PI Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;  
 PI Zhou P, Goodrich R, Drmanac RT;  
 XX  
 DR WPI: 2001-442253/47.  
 DR P-PSDB: AAM39772.  
 XX  
 PT Novel nucleic acids and polypeptides, useful for treating disorders such  
 PT as central nervous system injuries.  
 XX  
 PS Claim 1; SEQ ID NO 1131; 10078bp; English.  
 XX  
 CC The invention relates to human nucleic acids (AA157798-AA161369) and the  
 CC encoded polypeptides (AAM38642-AAM42213) with neurotropic,  
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
 CC in gene therapy. A composition containing a polypeptide or polynucleotide  
 CC of the invention may be used to treat diseases of the peripheral nervous  
 CC system, such as peripheral nervous injuries, peripheral neuropathy and  
 CC localised neuropathies and central nervous system diseases, such as  
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
 CC utilisation of the activities such as: Immune system suppression,  
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
 CC assays for receptor activity, arthritis and inflammation, leukemias and  
 CC C.N.S disorders. Note: The sequence data for this patent did not form  
 CC part of the printed specification  
 CC  
 SO Sequence 1767 BP; 393 A; 503 C; 493 G; 376 T; 0 U; 2 Other;  
 Query Match 25.9%; Score 832.2; DB 4; Length 1767;  
 Best Local Similarity 99.1%; Pred. No. 2.3e-229;  
 Matches 837; Conservative 0; Mismatches 8; Indels 0; Gaps 0;  
 QY 1 GAAAGGTTGCGAAGTGGCGCGCTTGAGCGAGAGAGCTGGACATGAATGAATATT 60  
 DB 139 GAAAGGTTGCGAAGTGGCGCGCTTGAGCGAGAGAGCTGGACATGAATGAATATT 198  
 QY 61 ACTGTTGCTGAACGTGCGCAAGGAGGCTTCTGAAGAGCTGAAGCTGCTACCGGA 120  
 DB 199 ACTGTTGCTGAACGTGCGCAAGGAGGCTTCTGAAGAGCTGAAGCTGCTACCGGA 258  
 QY 121 GGGCTGTATGCTCTACCATCAGACAGACAGACCCAGAGCTCAAGTCAACAGCGG 180  
 DB 259 GGGCTGTATGCTCTACCATCAGACAGACAGACCCAGAGCTCAAGTCAACAGCGG 318  
 QY 181 AAGAGCTGTTAAGCTGTTGACAGAGGCTTATGAAGGCTTATGAGCCCAACCAAGG 240  
 DB 319 AAGAGCTGTTAAGCTGTTGACAGAGGCTTATGAAGGCTTATGAGCCCAACCAAGG 378  
 QY 241 CCATCTATGATATATATGAGAGAGAGAGCTGAAATGAAAGATGGAGGTTTGAAA 300  
 DB 379 CCATCTATGATATATATGAGAGAGAGAGAGCTGAAATGAAAGATGGAGGTTTGAAA 438  
 QY 301 GAGAGAGAAACCCCTGCTGAAATTTGAGAGAGGTTTGAGCGGCTGACAGAGAGAGAG 360  
 DB 439 GAGAGAGAAACCCCTGCTGAAATTTGAGAGAGGTTTGAGCGGCTGACAGAGAGAGAG 498  
 QY 361 AGAGAGATTGTCAGACAGAGAAACAATCCAGAGAGAGATGAGGTTGAGTATATGCA 420

DB 499 AGAGAGATTGTCAGACAGAGAAACAATCCAGAGAGAGATGAGGTTGAGTATATGCA 558  
 QY 421 CCGACCTTTTGTATGCTATGATGAGAGATGAGATGATGTCGCGAGTAGTACCTTTCGC 480  
 DB 559 CCGACCTTTTGTATGCTATGATGAGAGATGAGATGATGTCGCGAGTAGTACCTTTCGC 618  
 QY 481 AGATTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 540  
 DB 619 AGATTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 678  
 QY 541 ACAAGCCTATCTCTCTGGAAGCCTCTCAACCCAGAAATGAAATGAGAGGTTTCCATTA 600  
 DB 679 ACAAGCCTATCTCTCTGGAAGCCTCTCAACCCAGAAATGAAATGAGAGGTTTCCATTA 738  
 QY 601 ACTTGGGCTCAGACGATGATCTTGGCAAGGAGATGGAGAGATTGGAATTTGAGCTG 660  
 DB 739 ACTTGGGCTCAGACGATGATCTTGGCAAGGAGATGGAGAGATTGGAATTTGAGCTG 798  
 QY 661 GAGACCTACAGGGGCTTGTGCTGATCAAGCTGTTCCGTATCTTCAACCAAGATGCT 720  
 DB 799 GAGACCTACAGGGGCTTGTGCTGATCAAGCTGTTCCGTATCTTCAACCAAGATGCT 858  
 QY 721 TTGTGACAAACAATGCTGCTCTGCAAGTTTTCATCCGTTGGAATCCGACCCGCTGACCA 780  
 DB 859 TTGTGACAAACAATGCTGCTCTGCAAGTTTTCATCCGTTGGAATCCGACCCGCTGACCA 918  
 QY 781 CTGTCTTACTGCTGGAACCTTGAACAAGAACCGTGGGCTTACCTGCAAGTGGAGTGA 840  
 DB 919 CTGTCTTACTGCTGGAACCTTGAACAAGAACCGTGGGCTTACCTGCAAGTGGAGTGA 978  
 QY 841 TCCAG 845  
 DB 979 CCCC 983  
 RESULT 12  
 ADB48910  
 ID ADB48910 standard; cDNA; 1767 BP.  
 AC ADB48910;  
 XX  
 DT 04-DEC-2003 (first entry)  
 XX  
 DE Novel human cDNA SEQ ID NO 820.  
 XX  
 KW sf; cancer; neurodegenerative disease; human.  
 XX  
 OS Homo sapiens.  
 XX  
 FN US2003104529-A1.  
 XX  
 PD 05-JUN-2003.  
 XX  
 PF 04-JAN-2002; 2002US-00037270.  
 XX  
 PR 21-JAN-2000; 2000US-00488725.  
 PR 25-APR-2000; 2000US-00552317.  
 PR 19-JUL-2000; 2000US-00620312.  
 XX  
 PA (ZHOU/) ZHOU P.  
 PA (TRANG/) TRANG Y T.  
 PA (LIUC/) LIU C.  
 PA (ASUN/) ASUNDI V.  
 PA (DRMA/) DRMANAC R T.  
 XX  
 PI Zhou P, Tang YT, Liu C, Neundi V, Drmanac RT;  
 XX  
 DR WPI: 2003-678194/64.  
 DR  
 XX  
 PT New polynucleotide, useful for treating diseases e.g., cancer or  
 PT neurodegenerative diseases.



```
Db 13 GAAAGTTGCGAAGATGCGACGCGCTTGACGAGAGAGAGCTGGAACAATGAGACTATT 72
Qy 61 ACTGTTGCTGAAGCTGCGAGGAGGCTCTTCTGAAGACTGAAAGCTGCTTACCGGA 120
Db 73 ACTGTTGCTGAAGCTGCGAGGAGGCTCTTCTGAAGACTGAAAGCTGCTTACCGGA 132
Qy 121 GGCTCTGATGCTCTACCATCCAGACAGACAGACCCAGAGCTCAAGTCAAGGCGG 180
Db 133 GGCTCTGATGCTCTACCATCCAGACAGACAGACCCAGAGCTCAAGTCAAGGCGG 192
Qy 181 AACGACTGTTAACTGTTTCAACGAGCTTATGAAAGTCTTATGAGACCCCAACAGGG 240
Db 193 AACGACTGTTAACTGTTTCAACGAGCTTATGAAAGTCTTATGAGACCCCAACAGGG 252
Qy 241 CCATCTATGATATATATATGGAAGAGAGACTGGAATATGAAAGATGGAGGTTGTGAAA 300
Db 253 CCATCTATGATATATATATGGAAGAGAGACTGGAATATGAAAGATGGAGGTTGTGAAA 312
Qy 301 GGAGGAGAACCCCTGCG-TGAAATTGAGAGAGATTGAGCGGCTGACAGAGAGAGAA 359
Db 313 GGAGGAGAACCCCTGCGTGGAAATTCGAGAGAGATTGAGCGGCTGAGAGAGAGAA 372
Qy 360 GAGAGAGATTGACAGCAGCAGCAACCAATCCAAAGGAGACATCAGCGTTGAGTAGATCC 419
Db 373 GAGAGAGATTGACAGCAGCAGCAACCAATCCAAAGGAGACATCAGCGTTGAGTAGATCC 432
Qy 420 ACCGACCTTTTGTATCGCTATGATGAGAGATGAAAGTGTCCGCGAGTAGCTTTCCG 479
Db 433 ACCGACCTTTTGTATCGCTATGATGAGAGATGAAAGTGTCCGCGAGTAGCTTTCCG 492
Qy 480 CAGATTGAAATTAATTAATAATGACATATCCAGTCCATTGAGGACCCCTTGACAGCGCA 539
Db 493 CAGATTGAAATTAATTAATAATGACATATCCAGTCCATTGAGGACCCCTTGACAGCGCA 552
Qy 540 GACACAGCCATCTCTCTGGAAGCTCTCAACCCAGATGGAATGAGAGAGTTCATTT 599
Db 553 GACACAGCCATCTCTCTGGAAGCTCTCAACCCAGATGGAATGAGAGAGTTCATTT 612
Qy 600 AACTTTGGGCTCAGACGAGTAATCTTCCGCAAGGATGGAGAGTGGAAATTTGAGCT 659
Db 613 AACTTTGGGCTCAGACGAGTAATCTTCCGCAAGGATGGAGAGTGGAAATTTGAGCT 672
Qy 660 GAGAGACCTACAGGGGCTTTGTTGGGCTCAAGGCTTCGTAATCTCACACCAAGATGC 719
Db 673 GAGAGACCTACAGGGGCTTTGTTGGGCTCAAGGCTTCGTAATCTCACACCAAGATGC 732
Qy 720 TTTGTGACAAACAACTGTGCTCTGCAAGTTTTCATCCCGTGAATCCGACCCGCGCTGAC 779
Db 733 TTTGTGACAAACAACTGTGCTCTGCAAGTTTTCATCCCGTGAATCCGACCCGCGCTGAC 792
Qy 780 ACTGTCTAGCTCGGAACCTTGAACAAGAACCCGTGGGCTTACTGCAAGTGGCGATGGGGT 839
Db 793 ACTGTCTAGCTCGGAACCTTGAACAAGAACCCGTGGGCTTACTGCAAGTGGCGATGGGGT 852
Qy 840 ATCCAG 845
Db 853 TCCCCG 858
```

## RESULT 14

AAS67386 standard; cDNA; 1098 BP.

AAS67386;

13-FEB-2002 (first entry)

DNA encoding novel human diagnostic protein #3190.

Human; chromosome mapping; gene mapping; gene therapy; forensic;  
food supplement; medical imaging; diagnostic; genetic disorder; ss.  
Homo sapiens.

```
XX 11-0CT-2001.
XX 3-0-MAR-2001; 2001WO-US008631.
XX 31-MAR-2000; 2000US-00540217.
XX 23-AUG-2000; 2000US-00649167.
XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
XX P-PSDB; ABG03199.
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensic, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity.
XX Claim 1; SEQ ID NO 3190; 103bp; English.
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
XX sequences. (I) is useful as hybridisation probes, polymerase chain
XX reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
XX and in recombinant production of (II). The polynucleotides are also used
XX in diagnostics as expressed sequence tags for identifying expressed
XX genes. (I) is useful in gene therapy techniques to restore normal
XX activity of (II) or to treat disease states involving (II). (II) is
XX useful for generating antibodies against it, detecting or quantitating a
XX polypeptide in tissue, as molecular weight markers and as a food
XX supplement. (II) and its binding partners are useful in medical imaging
XX of sites expressing (II). (I) and (II) are useful for treating disorders
XX involving aberrant protein expression or biological activity. The
XX polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensic, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
XX coding sequences of the invention. Note: The sequence data for this
XX patent did not appear in the printed specification, but was obtained in
XX electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 1098 BP; 271 A; 281 C; 320 G; 226 T; 0 U; 0 Other;
XX
XX Query Match 25.5%; Score 819.6; DB 5; Length 1098;
XX Best Local Similarity 98.8%; Pred. No. 7,7e-226;
XX Matches 836; Conservative 0; Mismatches 9; Indels 1; Gaps 1;
```





**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 13, 2004, 08:52:48 ; Search time 218 Seconds  
(Without alignments)  
8179.162 Million cell updates/sec

Title: US-09-787-678A-12

Perfect score: 3213  
Sequence: 1 gaaagctgcgaagatgCG.....gaaagcagaaaaaaa 3213

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

1: Issued Patente NA: \*  
2: /cgn2\_6/ptodata/2/ina/5A COMB.seq: \*  
3: /cgn2\_6/ptodata/2/ina/5B COMB.seq: \*  
4: /cgn2\_6/ptodata/2/ina/6A COMB.seq: \*  
5: /cgn2\_6/ptodata/2/ina/6B COMB.seq: \*  
6: /cgn2\_6/ptodata/2/ina/PTUS COMB.seq: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	832.2	25.9	1767	4	US-09-620-312D-820
2	455	14.2	455	4	US-09-404-879A-326
3	83.2	2.6	1261	4	US-09-724-864-34
4	58	1.8	1330	2	US-08-868-288A-4
5	58	1.8	1330	3	US-09-235-373-4
6	58	1.8	1330	3	US-09-388-993-4
7	57.2	1.8	2349	2	US-08-974-546-2
8	56.8	1.8	7218	1	US-08-233-463-14
9	48.4	1.5	1308	4	US-09-023-655-883
10	48.4	1.5	1318	4	US-09-665-479A-7
11	48.4	1.5	2607	4	US-09-976-594-226
12	44.2	1.4	713	4	US-09-023-655-300
13	42	1.3	297	4	US-09-313-294A-6215
14	42	1.3	1700	2	US-08-897-340-4
15	42	1.3	1700	2	US-09-252-329-4
16	42	1.3	3748	2	US-08-686-417-2
17	41	1.3	505	4	US-09-621-976-15639
18	40.4	1.3	284	4	US-09-313-294A-1936
19	40.4	1.3	1756	2	US-08-879-260-3
20	40.4	1.3	1811	3	US-09-231-529-5
21	40.4	1.3	1811	3	US-08-977-816-5
22	40.2	1.3	635	4	US-09-669-751-166
23	40.2	1.3	4403765	3	US-09-103-840A-2
24	40.2	1.3	4411529	3	US-09-103-840A-1
25	39.2	1.2	1166	4	US-09-072-596-323
26	39.2	1.2	1166	4	US-09-072-967-328
27	38.8	1.2	364	4	US-09-621-976-17202

28	38.6	1.2	1482	4	US-09-613-303-18	Sequence 18, Appl
29	38.6	1.2	1482	4	US-10-267-311-18	Sequence 18, Appl
30	38.6	1.2	16442	3	US-08-781-891-208	Sequence 208, App
31	38.6	1.2	16442	4	US-09-618-166-208	Sequence 208, App
32	38.2	1.2	1926	4	US-09-249-585A-4	Sequence 4, Appl1
33	38.2	1.2	1931	2	US-09-130-114-2	Sequence 2, Appl1
34	38.2	1.2	2019	3	US-09-063-950-3	Sequence 3, Appl1
35	38.2	1.2	2555	4	US-09-866-028-68	Sequence 68, Appl
36	38.2	1.2	2852	3	US-09-063-950-1	Sequence 1, Appl1
37	38.2	1.2	6336	4	US-08-956-171E-219	Sequence 219, App
38	37.8	1.2	882	4	US-09-252-991A-2568	Sequence 2568, App
39	37.8	1.2	993	4	US-09-252-991A-2568	Sequence 2427, App
40	37.6	1.2	2481	4	US-09-894-998A-35	Sequence 35, Appl
41	37.6	1.2	5919	4	US-08-875-435B-2	Sequence 2, Appl1
42	37.4	1.2	505	4	US-09-621-976-15639	Sequence 15639, A
43	37.4	1.2	1755	1	US-08-068-395A-2	Sequence 2, Appl1
44	37.4	1.2	1755	1	US-08-464-365-2	Sequence 2, Appl1
45	37.4	1.2	2568	4	US-09-489-039A-5258	Sequence 5258, Ap

#### ALIGNMENTS

RESULT 1

US-09-620-312D-820

Sequence 820, Application US/09620312D

Patent No. 6569662

GENERAL INFORMATION:

APPLICANT: Tang, Y. Tom

APPLICANT: Liu, Chenghua

APPLICANT: Asundi, Vinod

APPLICANT: Zhang, Jie

APPLICANT: Ren, Feiyan

APPLICANT: Zhao, Rui-hong

APPLICANT: Zhao, Qing A.

APPLICANT: Wehrman, Tom

APPLICANT: Xue, Aidong J.

APPLICANT: Yang, Yonghong

APPLICANT: Wang, Jian-Rui

APPLICANT: Zhou, Ping

APPLICANT: Ma, Yundong

APPLICANT: Wang, Dunrui

APPLICANT: Wang, Zhiwei

APPLICANT: Drmanac, Radoje T.

TITLE OF INVENTION: No. 6569662e1 Nucleic Acids and

FILE REFERENCE: 784CIP2B

CURRENT FILING DATE: 2000-07-19

PRIOR APPLICATION NUMBER: 09/552,317

PRIOR FILING DATE: 2000-04-25

PRIOR APPLICATION NUMBER: 09/488,725

NUMBER OF SEQ ID NOS: 1105

SOFTWARE: pc\_files Version 1.0

SEQ ID NO 820

LENGTH: 1767

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: CDS

LOCATION: (153)..(1223)

FEATURE:

NAME/KEY: misc feature

LOCATION: (1)..(1767)

OTHER INFORMATION: n = a,t,c or g

US-09-620-312D-820

Query Match

Best Local Similarity

Matches 837; Conservative

25.9%; Score 832.2; DB 4;

Pred. No. 1,5e-227;

Indels 0; Gaps 0;

```
QY      1 GAAAGTTGGGAATGCGACGCGCTTGAGCGAGGAGAGTGA CAATGAAGACTATT 60
      |||
      139 GAAGGTGGGAATGCGACGCGCTTGAGCGAGGAGAGTGA CAATGAAGACTATT 198
QY      61 ACTGTTGCTGAACTGTCGAGGAGGAGGCTTTCTGAGAGCTGAAAGCTGCTTACCGA 120
      |||
      199 ACTGTTGCTGAACTGTCGAGGAGGAGGCTTTCTGAGAGCTGAAAGCTGCTTACCGA 258
QY      121 GGGCTGTATGCTCTACCTCAATCCAGACAGACAGAGCCCAAGAGCTCAAGTCAAGGCGG 180
      |||
      259 GGGCTGTATGCTCTACCTCAATCCAGACAGACAGAGCCCAAGAGCTCAAGTCAAGGCGG 318
QY      181 AAGCACTGTTAACTGTTCAACAGGCTTATGAGTGTCTAGTGA CCCCCTCAACGAGG 240
      |||
      319 AAGCACTGTTAACTGTTCAACAGGCTTATGAGTGTCTAGTGA CCCCCTCAACGAGG 378
QY      241 CCACTATGATATATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
      |||
      379 CCACTATGATATATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 438
QY      301 GGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
      |||
      439 GGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 498
QY      361 AGAGAGAGATTCAGACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
      |||
      499 AGAGAGAGATTCAGACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 558
QY      421 CCGAAGCTTTTGAATGCTATGATGAGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
      |||
      559 CCGAAGCTTTTGAATGCTATGATGAGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 618
QY      481 AGATGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 540
      |||
      619 AGATGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 678
QY      541 ACAAGAGCATCTCTCTGAGAGCTCTCAACCCAGATGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
      |||
      679 ACAAGAGCATCTCTCTGAGAGCTCTCAACCCAGATGAGAGAGAGAGAGAGAGAGAGAGAGAG 738
QY      601 ACTTTGGCTGACAGAGATTAATTTGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
      |||
      739 ACTTTGGCTGACAGAGATTAATTTGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 798
QY      661 GAGAGCTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
      |||
      799 GAGAGCTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 858
QY      721 TTGTGACAAACAACTGTCTCTGAGATTTTCAATCCGTGAGATCCGACCCGCTGACCA 780
      |||
      859 TTGTGACAAACAACTGTCTCTGAGATTTTCAATCCGTGAGATCCGACCCGCTGACCA 918
QY      781 CTGTCTTACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840
      |||
      919 CTGTCTTACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 978
QY      841 TCCAG 845
      |||
      979 CCCCC 983
```

```
RESULT 2
US-09-404-879A-326/c
; Sequence 326, Application US/09404879A
; Patent No. 6468546
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: King, Gordon E.
; APPLICANT: Algate, Paul A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.462C2
; CURRENT APPLICATION NUMBER: US/09/404,879A
```

```
; CURRENT FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 393
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 326
; LENGTH: 455
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-404-879A-326
```

```
Query Match      14.2%; Score 455; DB 4; Length 455;
Best Local Similarity 100.0%; Pred. No. 5,1e-120; Indels 0; Gaps 0;
Matches 455; Conservative 0; Mismatches 0;
```

```
QY      1074 AAGGTGAGAGTTCACAGGAGCGTTTCTCTCAAGTCAAGCTCAACAGGAGCAGTGAACA 1133
      |||
      455 AAGGTGAGAGTTCACAGGAGCGTTTCTCTCAAGTCAAGCTCAACAGGAGCAGTGAACA 396
QY      1134 TACTTCTTCCATTAATCACTTGAACGAGACAGCTTTCGCCAGCGCATGTTCTATGCGACC 1193
      |||
      395 TACTTCTTCCATTAATCACTTGAACGAGACAGCTTTCGCCAGCGCATGTTCTATGCGACC 336
QY      1194 GTGGGAGCTCTAGTGTCTACTTTCGATGACACCGTGTATCATCAACCATACCTCAAG 1253
      |||
      335 GTGGGAGCTCTAGTGTCTACTTTCGATGACACCGTGTATCATCAACCATACCTCAAG 276
QY      1254 GCTCAAGAAAGAGAGAGATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1313
      |||
      275 GCTCAAGAAAGAGAGAGATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 216
QY      1314 AAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1373
      |||
      215 AAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 156
QY      1374 GAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1433
      |||
      155 GAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 96
QY      1434 AATGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1493
      |||
      95 AATGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 36
QY      1494 TGCTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1528
      |||
      35 TGCTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1
```

```
RESULT 3
US-09-724-864-34
; Sequence 34, Application US/09724864
; Patent No. 6380362
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Morrison, James G.
; TITLE OF INVENTION: Polynucleotides, polypeptides expressed
; FILE REFERENCE: 11000.1050U1
; CURRENT APPLICATION NUMBER: US/09/724,864
; PRIOR FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: U.S. No. 6380362 60/171,678
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 34
; LENGTH: 1261
; TYPE: DNA
; ORGANISM: Mouse
US-09-724-864-34
```

```
Query Match      2.6%; Score 83.2; DB 4; Length 1261;
Best Local Similarity 55.8%; Pred. No. 2e-13; Indels 72; Gaps 10;
Matches 404; Conservative 0; Mismatches 248;
```

```
QY      1703 GAACAGATTTTAAAGGCGCAAAAATCTTTCTGAGAGCTACAAATTGGAA-A 1761
```



```

Db      24      GAACTTCGAGGAAAAAGAGGCTACGAAAACTTTTCTTAAGAGGCTTACAAATTGGAAAGC 83
Oy      1762     TGAAAAAAACCAGACATCAGATGTTTTATTTATTTATTAATTAAGAGGTGTACCA 1821
Db      84      AGGGAAGAACCCGACATGATGATGTTTTAGTTATTTCTCCAGAA-----GGGGGCACT 137
Oy      1822     TTATCAATTATATGTGAAGGACATGCAGACACCCAGCTTTTGAGGGTCTGGGGGTAGA 1881
Db      138     GTATCAATTATATGTGAAGGACATGCAGACAGCTGTGCTCATGTGGCTGT-GGGGTAAGA 196
Oy      1882     CTGAGGACGACCCCACTGGGAACCAAGACTGCAGCTGGGCCCATGCTGTTTTCCCAAGAT 1941
Db      197     CTGAGGAGACCCCTCTGGGCCAGACCCCAAGACGGCCATG-----TCTCTCCAGAGATC 248
Oy      1942     CAGTTCTCGAAGGAAAGGGCTCTGGCCCTGACCTCCGCTGTGTCCGAGACACAGTGTGA 2001
Db      249     ATGTTCTGAGAGGTCAAGCCCTGTGCTT----- 278
Oy      2002     CCGACAGCCCGCCGCTGTAGTCTTGTGGCTGGGTCTGAGAGGTGTCTGTGAGCACCTGC 2061
Db      279     -----CTATTAAGTGGCTGTGTGCACAGCAGCTCTGTGAGGTATTTGAAACATTCGTC 330
Oy      2062     CCTCACACAGAGAGCGTGAAGCACTTCTGCAGTCCAGCTGAACATGGGAAACAACTGA 2121
Db      331     TGT---CAACAATGGGACTGCTCTTCTGAAGCCCAAGCTGTTCTGGGAAACATGGGA 387
Oy      2122     AAAGCAGCAGGCTCTCCGCGTCAAGGAGCCTCTGCTGTGTGCTGCTTCCATGA-CCACT 2180
Db      388     GAAAGGAAGAGGTGTGTG-----TGTGTCTAGTAGTACCTTCCCAAGAGCACT 437
Oy      2181     CCTCTGCTGAATATTAATCTGCTGAATCTGGAGCAGATTTGGGTTATATAAATGCTT 2240
Db      438     CTCCTCTTCTGAAACCTCACTGAATGA-CTGGAGAGGACTCGGGGTTATATAAATGCTT 496
Oy      2241     TTATCTGAGAAACAAGGGTTTGGAAATTAAGTGTCTTTTTTCCCACTCCAGAGCTG 2300
Db      497     TTATCTGAGAAACAATGGGTTTGGAAACTAGTCTCTTTTCTTCCACTTTTACAGAGCTT 556
Oy      2301     CTCAAGTCATTCCACCGGCCCCCTTGCGCTTGGGACAGGGTATGTAACTCCCAATCCAG 2368
Db      557     CTCGAATTAATCTCTCAGGCCCCCTGACTTGGACAGGTAGGGGGGAGACCTCGGGTCCAA 616
Oy      2361     GAGCTAGCCCTGCACAGAGTGGCTTCCCGTATCCCGGTGGGAAACGCGCTGCACAGC 2420
Db      617     GG---TGACGTGTCCAGGCAACACTGCCACATTTGCTAAGAGAAAGGCCCTGTGCCAGT 673
Oy      2421     GGGC 2424
Db      674     GGAC 677

RESULT 4
US-08-868-288A-4
; Sequence 4, Application US/08868288A
; Patent No. 5922567
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; APPLICANT: Lal, Preeti
; APPLICANT: Bandman, Olga
; TITLE OF INVENTION: TWO NEW HUMAN DNAJ-LIKE PROTEINS
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS

```

```

SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/868,288A
FILING DATE: June 3, 1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: P-0309 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1330 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: HMT2RAT01
CLONE: 260873
US-08-868-288A-4

Query Match 1.8%; Score 58; DB 2; Length 1330;
Best Local Similarity 55.1%; Pred. No. 3.3e-06;
Matches 134; Conservative 0; Mismatches 103; Indels 6; Gaps 1;

QY 46 ACATGGAAGATATTATCTCTGTTGCTGAACGTGCCGAGGAGGCTCTTTCTGAAGACTGA 105
DB 155 ACATGGTGGATATTACTATGAACTTTAGGCGTGACAGAGCATGCTTCAACCCAGATATTA 214
QY 106 AAGTCCTTACCGGAGGCTGTGTNGCTCTACATCCAGACAGACAGACAGACCCAGAC 165
DB 215 AAAAGGCATATCGAAACTGCACTGAAGTGGCATTCAGATA-----AAATCTGAGA 268
QY 166 TCAAGTCAAGGCGGAGGAGTGTGTTAACTTTGTTCAACGAGCTTATGAAGTGTAGTG 225
DB 269 ATAAAGAAAGACGAGAGAAATTCAAGCAAGTAGCGAGGCGATATGAAGTGTGTGG 328
QY 226 ACCCCCAACCAAGGCGCATCTATATATATATGGAAGAGAGACTGGAATGAGAGAT 285
DB 329 ATGTGTAAGAAACGGGACATCTATGACAAATATGGCAAAAGAGATTAATGTGTGGNG 388
QY 286 GGG 288
DB 389 GNG 391

RESULT 5
US-09-235-373-4
Sequence 4, Application US/09235373
Patent No. 6001598
GENERAL INFORMATION:
APPLICANT: Au-Young, Janice
APPLICANT: Lal, Preeti
APPLICANT: Bandman, Olga
TITLE OF INVENTION: TWO NEW HUMAN DNA-LIKE PROTEINS
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESS: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:

```

APPLICATION NUMBER: US/09/235,373  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/868,288  
FILING DATE: June 3, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0309 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1330 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: HNT2RAT01  
CLONE: 260873  
US-09-235-373-4

Query Match 1.8%; Score 58; DB 3; Length 1330;  
Best Local Similarity 55.1%; Pred. No. 3.3e-06;  
Matches 134; Conservative 0; Mismatches 103; Indels 6; Gaps 1;

QY 46 ACAATGAAGACTATTCTGCTGTGAGAGCGGAGGCGCTCTTCTGAAGAGCTGA 105  
DB 155 ACATGGGATTTACTATGAAAGTTCTAGCGCTGAGAGACATGCTCAACCGAGGATATTA 214  
QY 106 AAGCTGCTTACCGGAGGCTCTGTATGCTTACCATCCAGACAGACAGACAGACCCAGAGC 165  
DB 215 AAAGGCAATATCGGAAACTGGCACTGAAGTGGCATCCAGATTA-----AAATCTGAGA 268  
QY 166 TCAAGTCAAGCGGAGGAGCACTGTTTAACTTGTTCACCGCTTATGAAGTCTTAGTG 225  
DB 269 ATAAAGAAAGAGAGAGAGAAATTCAGCAAGTAGCGGAGCATATGAAGTCTGTCCG 328  
QY 226 ACCCCCAACCGAGGCGCATATGATATATATGAGAAAGAGAGAGCTGGAATGGAAGAT 285  
DB 329 ATGCTAAGAAACGGAGCATCTATGACAAATATGCGCAAGAGATTAAATGTTGGGNGG 388  
QY 286 GGG 288  
DB 389 GNG 391

RESULT 6  
US-09-388-993-4  
Sequence 4, Application US/09388993  
Patent No. 6043222  
GENERAL INFORMATION:  
APPLICANT: Au-Young, Janice  
APPLICANT: Lal, Preeti  
APPLICANT: Bandman, Olga  
TITLE OF INVENTION: TWO NEW HUMAN DNA-LIKE PROTEINS  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/388,993  
FILING DATE:

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/868,288  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0309 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1330 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: HNT2RAT01  
CLONE: 260873  
US-09-388-993-4

Query Match 1.8%; Score 58; DB 3; Length 1330;  
Best Local Similarity 55.1%; Pred. No. 3.3e-06;  
Matches 134; Conservative 0; Mismatches 103; Indels 6; Gaps 1;

QY 46 ACAATGAAGACTATTCTGCTGTGAGAGCGGAGGCGCTCTTCTGAAGAGCTGA 105  
DB 155 ACATGGGATTTACTATGAAAGTTCTAGCGCTGAGAGACATGCTCAACCGAGGATATTA 214  
QY 106 AAGCTGCTTACCGGAGGCTCTGTATGCTTACCATCCAGACAGACAGACAGACCCAGAGC 165  
DB 215 AAAGGCAATATCGGAAACTGGCACTGAAGTGGCATCCAGATTA-----AAATCTGAGA 268  
QY 166 TCAAGTCAAGCGGAGGAGCACTGTTTAACTTGTTCACCGCTTATGAAGTCTTAGTG 225  
DB 269 ATAAAGAAAGAGAGAGAAATTCAGCAAGTAGCGGAGCATATGAAGTCTGTCCG 328  
QY 226 ACCCCCAACCGAGGCGCATATGATATATATGAGAAAGAGAGAGCTGGAATGGAAGAT 285  
DB 329 ATGCTAAGAAACGGAGCATCTATGACAAATATGCGCAAGAGATTAAATGTTGGGNGG 388  
QY 286 GGG 288  
DB 389 GNG 391

RESULT 7  
US-08-974-546-2  
Sequence 2, Application US/08974546  
Patent No. 5945287  
GENERAL INFORMATION:  
APPLICANT: Hillman, Jennifer L.  
APPLICANT: Lal, Preeti  
APPLICANT: Corley, Neil C.  
APPLICANT: Shah, Purvi  
TITLE OF INVENTION: TWO HUMAN HEAT SHOCK PROTEIN HOMOLOGS  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Dr.  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/974,546  
FILING DATE: Filed Herewith  
PRIOR APPLICATION DATA:

```

: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Billings, Lucy J.
: REGISTRATION NUMBER: 36,749
: REFERENCE/DOCKET NUMBER: PF-0428
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 650-855-0555
: TELEFAX: 650-845-4166
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2349 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: IMMEDIATE SOURCE:
: LIBRARY: BRAITUT21
: CLONE: 2525691
:
US-08-974-546-2

Query Match          1.8%; Score 57.2; DB 2; Length 2349;
Best Local Similarity 56.9%; Pred. No. 7.8e-06;
Matches 132; Conservative 0; Mismatches 88; Indels 12; Gaps 1;

QY 52 AAGACTATTACTCTGCTGTAACGTGCGCAGGAGGCGCTTTGTAAGAGCTGAAGCTG 111
DB 109 AAGATTATTACAAATTTCTTGGATCCATCGGGGCCAAGAGATGATCAAGAAAG 168
QY 112 CTTACCGGAGGCTGTGATGCTCTACCATCCAGACAGACAGACCAGAGCTCAAGT 171
DB 169 CTTACCGGAGATGCGCTTGAAGTACCAACCAGACAAATTAAGAACCCAAAC----- 221
QY 172 CACAGCGCGAAGCACTGTTTAACTGTTCACAGGCTTAAAGTGTGTAAGCCCC 231
DB 222 -----GCTGAGAGAAAGTTAAGAGATTCAGAGGCGCTTAAGTGTGTAAGTACCCCA 276
QY 232 AAACGAGGCGCATGTATATATATATGGAAGAGAGCATGTGAATGGAAGG 283
DB 277 AGAAACGGGCGCTGTATGACCAATGAGGAGAAAGCGCTTGAAGCCGCGG 328

RESULT 8
US-08-232-463-14/c
: Sequence 14, Application US/08232463
: Patent No. 5670367
: GENERAL INFORMATION:
: APPLICANT: DORNER, F.
: APPLICANT: SCHEIFLINGER, F.
: APPLICANT: FALKNER, F. G.
: TITLE OF INVENTION: RECOMBINANT FOMLOX VIRUS
: NUMBER OF SEQUENCES: 52
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Foley & Lardner
: STREET: 1800 Diagonal Road, Suite 500
: CITY: Alexandria
: STATE: VA
: COUNTRY: USA
: ZIP: 22313-0299
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/232,463
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION NUMBER:
: APPLICATION NUMBER: US/07/935,313
: FILING DATE:
: APPLICATION NUMBER: EP 91 114 300.6
: FILING DATE: 26-AUG-1991
: ATTORNEY/AGENT INFORMATION:
```

```

: NAME: BENT, Stephen A.
: REGISTRATION NUMBER: 29,768
: REFERENCE/DOCKET NUMBER: 30472/114 IMMU
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703) 836-9300
: TELEFAX: (703) 683-4109
: TELEX: 899149
: INFORMATION FOR SEQ ID NO: 14:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 7218 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: IMMEDIATE SOURCE:
: CLONE: PTZ9C-F15
:
US-08-232-463-14

Query Match          1.8%; Score 56.8; DB 1; Length 7218;
Best Local Similarity 4.1%; Pred. No. 2e-05;
Matches 16; Conservative 220; Mismatches 152; Indels 0; Gaps 0;

QY 143 AGACMACACAGAGACCCAGAGCTCAATGACAGAGCGAAGCACTGTTAACTGTGTA 202
DB 1421 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1362
QY 203 CCAGCTTATGAAGTGTGTAAGTACCCCAACAGGCGCATCTATGATATATAGGAA 262
DB 1361 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1302
QY 263 GAGAGCATGAATGAGATGAGAGTGTGGAAGAGAGAACCCCTGCTGAAT 322
DB 1301 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1242
QY 323 TCAGAGAGAGTTCGAGCGGCTGACAGAGAGAGAGAGAGAGATTGACAGCGAAC 382
DB 1241 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1182
QY 383 CAATCCCAAGGAACGATGCGCTTGAGATGATGCCAGCACTTTTATGCTATGA 442
DB 1181 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1122
QY 443 TGAGAGTATGAAGATGTGCGGAGTAGCTTTCGACATTTGAATTAATAATGCA 502
DB 1121 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1062
QY 503 CATATCCCACTTCATTGAGCGACCTTG 530
DB 1061 AAGCTCCCTGACTGACGCCAAGCTTG 1034

RESULT 9
US-09-023-655-883
: Sequence 883, Application US/09023655
: Patent No. 6607879
: GENERAL INFORMATION:
: APPLICANT: Cocks, Benjamin G.
: APPLICANT: Susan G. Stuart
: APPLICANT: Jeffrey J. Selhammer
: TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL GENE
: NUMBER OF SEQUENCES: 1508
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
: STREET: 3174 PORTER DRIVE
: CITY: PALO ALTO
: STATE: CALIFORNIA
: COUNTRY: USA
: ZIP: 94304
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Word Perfect 6.1 for windows/MS-DOS 6.2
```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/023,655
; FILING DATE: HEREWITH
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Zeller, Karen J.
; REGISTRATION NUMBER: 37,071
; REFERENCE/DOCKET NUMBER: PA-0001 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 883:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1308 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GENBANK
; CLONE: g1127832
; US-09-023-655-883

```

```

Query Match      1.5%; Score 48.4; DB 4; Length 1308;
Best Local Similarity 52.7%; Pred. No. 0.0018;
Matches 137; Conservative 0; Mismatches 111; Indels 12; Gaps 1;

```

```

QY 52 AAGACTATTACTGCTGTTGCTGAACGTGCGCAGGAGGCGCTTCTTGAAGAGCTGAAAGCTG 111
   |||||
DB 184 AAGACTATTACTGCTGTTGCTGAACGTGCGCAGGAGGCGCTTCTTGAAGAGCTGAAAGAG 243
   |||||
QY 112 CCAACCGGAGGCTCTGTATGCTCTCAATCCAGACAAGACAAGAGCCGAGACTCAAGT 171
   |||||
DB 244 CTTACCGAAGAAACAAGCCCTCAAAATTTATCCGACAAGACAAGAAATCTC----- 291
   |||||
QY 172 CACAGCGGAGGAGCACTGTTTAACTTTCACCAAGGCTTATGAAAGTCTTGAAGCTGACCC 231
   |||||
DB 292 CTCAGGAGAGGAGAAATTTAAAGAGGTCCAGAGCTTATGAGTATGAGTATCTTA 351
   |||||
QY 232 AAACCGGAGGCTCTGTATGATATATATGAGGAGAGAGACTGGAATGGAAGATGGAGG 291
   |||||
DB 352 AAAAGAGAGAAATATATGATCAGTTTGGGAGAGAGGTTGAAGAGAGGAGAGGTA 411
   |||||
QY 292 TTGTGGAAGAGAGAGAAC 311
   |||||
DB 412 CTGATGACAAGAGAGTACC 431
   |||||

```

## RESULT 10

US-09-665-479A-7

```

; Sequence 7, Application US/09665479A
; Patent No. 6673570
; GENERAL INFORMATION:
; APPLICANT: Itoh, Suniko
; APPLICANT: Itoh, Sunumu
; APPLICANT: Heldin, Carl-Henrik
; APPLICANT: Iken-Dijke, Peter
; TITLE OF INVENTION: SWAD ASSOCIATING POLYPEPTIDES
; FILE REFERENCE: L00461,70096.US
; CURRENT APPLICATION NUMBER: US/09/665,479A
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US 60/154,846
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 7
; LENGTH: 1318
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:

```

```

; NAME/KEY: CDS
; LOCATION: (177)..(1190)
; US-09-665-479A-7

```

```

Query Match      1.5%; Score 48.4; DB 4; Length 1318;
Best Local Similarity 52.7%; Pred. No. 0.0018;
Matches 137; Conservative 0; Mismatches 111; Indels 12; Gaps 1;

```

```

QY 52 AAGACTATTACTGCTGTTGCTGAACGTGCGCAGGAGGCGCTTCTTGAAGAGCTGAAAGCTG 111
   |||||
DB 184 AAGACTATTACTGCTGTTTGGGAATTTGAGAAAGAGCTTCAATGAAATATTTAAAGG 243
   |||||
QY 112 CCAACCGGAGGCTCTGTATGCTCTCAATCCAGACAAGACAAGAGCCGAGACTCAAGT 171
   |||||
DB 244 CTTACCGAAGAAACAAGCCCTCAAAATTTATCCGACAAGACAAGAAATCTC----- 291
   |||||
QY 172 CACAGCGGAGAGCACTGTTTAACTTTCACCAAGGCTTATGAAAGTCTTGAAGCTGACCC 231
   |||||
DB 292 CTCAGGAGAGGAGAAATTTAAAGAGGTCCAGAAAGCTTATGAAATGAGTATCTTA 351
   |||||
QY 232 AAACCGGAGGCTCTGTATGATATATATGAGGAGAGAGACTGGAATGGAAGATGGAGG 291
   |||||
DB 352 AAAAGAGAGAAATATATGATCAGTTTGGGAGAGAGGTTGAAGAGAGAGAGGTA 411
   |||||
QY 292 TTGTGGAAGAGAGAGAAC 311
   |||||
DB 412 CTGATGACAAGAGAGTACC 431
   |||||

```

## RESULT 11

US-09-976-594-226

```

; Sequence 226, Application US/0976594
; Patent No. 6673549
; GENERAL INFORMATION:
; APPLICANT: Purness, Michael
; APPLICANT: Buchbinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 226
; LENGTH: 2607
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6673549 138709.5
; US-09-976-594-226

```

```

Query Match      1.5%; Score 48.4; DB 4; Length 2607;
Best Local Similarity 52.7%; Pred. No. 0.0027;
Matches 137; Conservative 0; Mismatches 111; Indels 12; Gaps 1;

```

```

QY 52 AAGACTATTACTGCTGTTGCTGAACGTGCGCAGGAGGCGCTTCTTGAAGAGCTGAAAGCTG 111
   |||||
DB 233 AAGACTATTACTGCTGTTTGGGAATTTGAGAAAGAGCTTCAATGAAATATTTAAAGG 292
   |||||
QY 112 CCAACCGGAGGCTCTGTATGCTCTCAATCCAGACAAGACAAGAGCCGAGACTCAAGT 171
   |||||
DB 293 CTTACCGAAGAAACAAGCCCTCAAAATTTATCCGACAAGACAAGAAATCTC----- 340
   |||||
QY 172 CACAGCGGAGAGCACTGTTTAACTTTCACCAAGGCTTATGAAAGTCTTGAAGCTGACCC 231
   |||||
DB 341 CTCAGGAGAGGAGAAATTTAAAGAGGTCCAGAAAGCTTATGAAATGAGTATCTTA 400
   |||||
QY 232 AAACCGGAGGCTCTGTATGATATATATGAGGAGAGAGACTGGAATGGAAGATGGAGG 291
   |||||
DB 401 AAAAGAGAGAAATATATGATCAGTTTGGGAGAGAGGTTGAAGAGAGAGAGGTA 460
   |||||

```

OY 292 TTGTGAAAGGAGGAGACC 311  
Db 461 CTGATGACAGAGAGGTRACC 480

## RESULT 12

US-09-023-655-300  
; Sequence 300, Application US/09023655  
; Patent No. 6607879  
; GENERAL INFORMATION:  
; APPLICANT: Cocks, Benjamin G.  
; APPLICANT: Susan G. Stuart  
; APPLICANT: Jeffrey J. Seilhamer  
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL GENE  
; NUMBER OF SEQUENCES: 1508  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
; STREET: 3174 PORTER DRIVE  
; CITY: PALO ALTO  
; STATE: CALIFORNIA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/023,655  
; FILING DATE: HEREMITH  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Zeller, Karen J.  
; REGISTRATION NUMBER: 37,071  
; REFERENCE/DOCKET NUMBER: PA-0001 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (650) 855-0555  
; TELEFAX: (650) 845-4166  
; INFORMATION FOR SEQ ID NO: 300:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 713 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: LATRUT02  
; CLONE: 1350210  
; US-09-023-655-300

Query Match 1.4%; Score 44.2; DB 4; Length 713;  
Best Local Similarity 61.9%; Pred. No. 0.02;

Matches 70; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

OY 50 TGAAGACTATTACTGCTGTGTAACGTGCGAGGAGGCTCTTCTGAAGAGTGAAGC 109  
Db 358 TAAAGATTACTATGATGATCCTGGGGGTAGCAGAGGGCTCGATGAGACTGAAGAA 417  
OY 110 TGCTACCGAGGCTCTGTATGCTCTTACATCCAGACAGACAGACAGACCCAG 162  
Db 418 GGCTACCGAGAGCTGGCCCTCAATTCCACCAGACCAAGACAGCAGCAGCCTG 470

## RESULT 13

US-09-313-294A-6215  
; Sequence 6215, Application US/09313294A  
; Patent No. 6476212  
; GENERAL INFORMATION:  
; APPLICANT: Laligudi, Raghunath V.

; APPLICANT: Ito, Laura Y.  
; APPLICANT: Sherman, Bradley K.  
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR  
; FILE REFERENCE: PL-0017 US  
; CURRENT APPLICATION NUMBER: US/09/313,294A  
; CURRENT FILING DATE: 1999-05-14  
; NUMBER OF SEQ ID NOS: 7600  
; SOFTWARE: PERL Program  
; SEQ ID NO 6215  
; LENGTH: 297  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURES:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Incyte ID No. 6476212 700351424H1  
; LOCATION: 4, 10-11, 27, 33-34, 39, 72-88, 114, 263, 274-280  
; OTHER INFORMATION: a, t, c, g, or other  
US-09-313-294A-6215

Query Match 1.3%; Score 42; DB 4; Length 297;  
Best Local Similarity 57.7%; Pred. No. 0.051;  
Matches 75; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

OY 35 GGAGAGCTGGAAGTGAAGACTATTACTGCTGTGTAAGTGGCGAGGAGGCTCTTC 94  
Db 121 GAAGAGAGCGACACCAAGTACTACGAGATCTCGGGTCCCAAGTGGGCTCCA 180  
OY 95 TGAAGAGCTGAAGCTGCTTACCGAGGCTCTGTATGCTTACCATCCAGACAGCAG 154  
Db 181 GAGAGACTCAAGAGGCTTACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 240  
OY 155 AGACCCAGAG 164  
Db 241 TGAACCCGAG 250

## RESULT 14

US-08-897-340-4  
; Sequence 4, Application US/08897340  
; Patent No. 5955306  
; GENERAL INFORMATION:  
; APPLICANT: Gimeno, Carlos J. and Errada, Patrick, R.  
; TITLE OF INVENTION: Weight Control Pathway Genes and Uses  
; NUMBER OF SEQUENCES: 36  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: LAHIVE & COCKFIELD, LLP  
; STREET: 28 State Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/897,340  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/715,032  
; FILING DATE: 17-SEP-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Silverl, Jean M.  
; REGISTRATION NUMBER: 39,030  
; REFERENCE/DOCKET NUMBER: MNT-005CP  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)227-7400  
; TELEFAX: (617)227-5941  
; INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:  
LENGTH: 1700 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-08-897-340-4

Query Match 1.3%; Score 42; DB 2; Length 1700;  
Best Local Similarity 59.0%; Pred. No. 0.14;  
Matches 72; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 35 GGAGAGCTGGAACAATGAAGACTATTACTGCTGTAACGTGCGAGAGGAGGCTCTTC 94  
DB 973 GAAGAAGACCAAGAGAAAGATTACTACAAGATCTGGAGTGGACAAGAAATGCTCTGA 1032  
QY 95 TGAAGAGCTGAAGAGCTGCTACCGGAGGCTCTGTATGCTTACCATCCAGACAAGACAG 154  
DB 1033 GGACGAGATCAAGAAAGCTTACCGAAACGGGCTTGATGACCATCCAGATCGGCACAG 1092  
QY 155 AG 156  
DB 1093 TG 1094

## RESULT 15

US-09-252-329-4  
Sequence 4, Application US/09252329  
Patent No. 6147192

## GENERAL INFORMATION:

APPLICANT: Gimeno, Carlos J. and Errada, Patrick, R.  
TITLE OF INVENTION: Weight Control Pathway Genes and Uses  
TITLE OF INVENTION: Therefore  
NUMBER OF SEQUENCES: 36  
CORRESPONDENCE ADDRESS:

ADDRESSEE: LAHYE & COCKFIELD, LLP

STREET: 28 State Street

CITY: Boston

STATE: Massachusetts

COUNTRY: USA

ZIP: 02109

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/252,329

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/897,340

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Silveri, Jean M.

REGISTRATION NUMBER: 39,030

REFERENCE/DOCKET NUMBER: MNI-005CP

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617)227-7400

TELEFAX: (617)227-5941

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 1700 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

US-09-252-329-4

QY 35 GGAGAGCTGGAACAATGAAGACTATTACTGCTGTAACGTGCGAGAGGAGGCTCTTC 94  
DB 973 GAAGAAGACCAAGAGAAAGATTACTACAAGATCTGGAGTGGACAAGAAATGCTCTGA 1032  
QY 95 TGAAGAGCTGAAGAGCTGCTACCGGAGGCTCTGTATGCTTACCATCCAGACAAGACAG 154  
DB 1033 GGACGAGATCAAGAAAGCTTACCGAAACGGGCTTGATGACCATCCAGATCGGCACAG 1092  
QY 155 AG 156  
DB 1093 TG 1094

Search completed: July 13, 2004, 14:54:04  
Job time : 227 secs

Query Match 1.3%; Score 42; DB 3; Length 1700;  
Best Local Similarity 59.0%; Pred. No. 0.14;  
Matches 72; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 13, 2004, 12:39:28 ; Search time 1501 Seconds  
(without alignments)  
10432.903 Million cell updates/sec

Title: US-09-787-678A-12

Perfect score: 3213  
Sequence: 1 gaaagcttggaagatgagc.....gaaagcagaaaaaaaaa 3213

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3183909 seqs, 2436941669 residues

Total number of hits satisfying chosen parameters: 6367818

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:\*

1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_PUB.seq:\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/1/pubpna/PTUTS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*  
10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq:\*  
12: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq:\*  
13: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*  
14: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*  
15: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*  
16: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq:\*  
17: /cgn2\_6/ptodata/1/pubpna/US10C\_NEW\_PUB.seq:\*  
18: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUBCOMB.seq:\*  
19: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3191.4	59.3	3230	15	US-10-000-897-32 Sequence 32, App1
2	1851.6	57.6	2454	16	US-10-104-047-1796 Sequence 1796, App
3	1608.8	50.1	3600	16	US-10-062-674-1802 Sequence 1802, App
4	1414.6	44.0	1454	16	US-10-264-049-601 Sequence 601, App
5	1413.6	44.0	1451	9	US-09-925-300-487 Sequence 487, App
6	1397.4	43.5	2052	16	US-10-108-260A-1143 Sequence 1143, App
7	1006.6	31.3	1014	15	US-10-000-897-31 Sequence 31, App1
8	832.2	25.9	1767	15	US-10-037-270-820 Sequence 820, App
9	832.2	25.9	1767	16	US-10-117-722-820 Sequence 820, App
10	456.8	14.2	844	9	US-09-927-738-16 Sequence 16, App1
11	455	14.2	455	9	US-09-884-441-326 Sequence 326, App
12	455	14.2	455	10	US-09-907-969-326 Sequence 326, App
13	455	14.2	455	10	US-09-827-271-326 Sequence 326, App
14	455	14.2	455	15	US-10-198-053-326 Sequence 326, App

15	452	14.1	479	9	US-09-764-868-167	Sequence 167, App
16	450.2	14.0	479	9	US-09-764-868-541	Sequence 541, App
17	448.6	14.0	502	10	US-09-918-995-19	Sequence 19, App1
18	427	13.3	486	10	US-09-918-995-622	Sequence 622, App
19	415.6	12.9	448	10	US-09-918-995-28082	Sequence 28082, A
20	306	9.5	821	17	US-10-416-907-54	Sequence 54, App1
21	302.6	9.4	455	9	US-09-954-531-422	Sequence 422, App
22	276.2	8.6	351	16	US-10-062-674-840	Sequence 840, App
23	271	8.4	355	15	US-10-000-897-33	Sequence 33, App1
24	121	3.8	188	9	US-09-783-590-7035	Sequence 7035, App
25	83.2	2.6	1261	10	US-09-866-050A-603	Sequence 603, App
26	83.2	2.6	1261	15	US-10-152-661-603	Sequence 603, App
27	72.4	2.3	495	10	US-09-918-995-32122	Sequence 32122, A
28	69	2.1	2248	17	US-10-437-963-29240	Sequence 29240, A
29	64	2.0	1363	13	US-10-425-114-12551	Sequence 12551, A
30	61.2	1.9	430	13	US-10-424-599-135287	Sequence 135287, A
31	60.8	1.9	538	13	US-10-424-599-90682	Sequence 90682, A
32	60.4	1.9	1106	13	US-10-425-114-16826	Sequence 16826, A
33	60	1.9	60	10	US-09-908-975-7874	Sequence 111, App
34	60	1.9	60	10	US-09-908-975-23265	Sequence 23265, A
35	60	1.9	392	9	US-09-864-761-4922	Sequence 4922, App
36	59.4	1.8	479	10	US-09-918-995-4804	Sequence 4804, App
37	59.4	1.8	1029	17	US-10-437-963-24371	Sequence 24371, A
38	59.4	1.8	1167	13	US-10-302-172-645	Sequence 645, App
39	59.4	1.8	1594	16	US-10-264-049-499	Sequence 499, App
40	59	1.8	744	13	US-10-424-599-49074	Sequence 49074, A
41	58.4	1.8	1339	13	US-10-424-599-135287	Sequence 135287, A
42	58	1.8	1177	17	US-10-437-963-88349	Sequence 88349, A
43	57.2	1.8	2458	15	US-10-252-157-133	Sequence 133, App
44	55.8	1.7	414	9	US-09-960-352-4372	Sequence 4372, App
45	55.2	1.7	1399	13	US-10-424-599-49073	Sequence 49073, A

ALIGNMENTS

RESULT 1  
US-10-000-897-32  
Sequence 32, Application US/10000897  
Publication No. US20030165852A1  
GENERAL INFORMATION:  
APPLICANT: Schueler, P.  
TITLE OF INVENTION: METHODS AND REAGENTS FOR IDENTIFYING RARE PETAL CELLS IN THE  
FILE REFERENCE: 11012-004-999  
CURRENT APPLICATION NUMBER: US/10/000,897  
CURRENT FILING DATE: 2001-11-15  
PRIOR APPLICATION NUMBER: 60/248,882  
PRIOR FILING DATE: 2000-11-15  
NUMBER OF SEQ ID NOS: 78  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 32  
LENGTH: 3230  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-000-897-32

Query Match	99.3%	Score 3191.4	DB 15	Length 3230	
Beat Local Similarity	99.8%	Pred. No. 0			
Matches 3206	Conservative	0	Mismatches 6	Indels 1	Gaps 1
Qy	1	GAAAGTTGGGAAGATGGCAAGCGCTTGAGCGAGAGAGCTGGACATGAAGACTATT	60		
Db	7	GAAAGTTGGGAAGATGGCAAGCGCTTGAGCGAGAGAGCTGGACATGAAGACTATT	66		
Qy	61	ACTCGTTGTAAGCGTGGAGGAGGCGCTTCTGAAGAGCTGAAGAGCTCTACCGGA	120		
Db	67	ACTCGTTGTAAGCGTGGAGGAGGCGCTTCTGAAGAGCTGAAGAGCTCTACCGGA	126		
Qy	121	GGCTGTATGCTCTACATCCAGACAGACAGACCCAGAGCTCAAGTCAAGCGG	180		
Db	127	GGCTGTATGCTCTACATCCAGACAGACAGACCCAGAGCTCAAGTCAAGCGG	186		

QY 181 AACGACTGTTAACTTGTTCACCGAGGCTTATGAAAGTGCTTAGTGAACCCCAACCAAGG 240  
| | | | |  
Db 187 AAGCACTGTTAACTTGTTCACCGAGGCTTATGAAAGTGCTTAGTGAACCCCAACCAAGG 246  
| | | | |  
QY 241 CCATCTATGATATATATATGGAAGAGAGACTGGAAATGGAAGATGGAGAGTTGTGAAA 300  
| | | | |  
Db 247 CCATCTATGATATATATATGGAAGAGAGACTGGAAATGGAAGATGGAGAGTTGTGAAA 306  
| | | | |  
QY 301 GAGAGAGAACCCCTGCTGAATTCGAGAGAGTTTGAGCGGCTGCAGAGAGAGAGAG 360  
| | | | |  
Db 307 GAGAGAGAACCCCTGCTGAATTCGAGAGAGTTTGAGCGGCTGCAGAGAGAGAGAG 366  
| | | | |  
QY 361 AGAGAGATTCGACGACGAGCAACCAATCCCAAGGAAAGATCAGCGTTGAGTATGCCA 420  
| | | | |  
Db 367 AGAGAGATTCGACGACGAGCAACCAATCCCAAGGAAAGATCAGCGTTGAGTATGCCA 426  
| | | | |  
QY 421 CCGACCTTTTGTATCGGTATGATGAGAGATGAAAGTGTCCGGCAGTACGTTCCGC 480  
| | | | |  
Db 427 CCGACCTTTTGTATCGGTATGATGAGAGATGAAAGTGTCCGGCAGTACGTTCCGC 486  
| | | | |  
QY 481 AGATTGAATTTAATTAATGCAATATCCAGTCCATTGAGGACCCCTTGACGCGCAG 540  
| | | | |  
Db 487 AGATTGAATTTAATTAATGCAATATCCAGTCCATTGAGGACCCCTTGACGCGCAG 546  
| | | | |  
QY 541 ACAACGCAATCTCTCTGGAAGCTCTCAAACCAAGATGGAATGGAAGAGTTCCATTA 600  
| | | | |  
Db 547 ACAACGCAATCTCTCTGGAAGCTCTCAAACCAAGATGGAATGGAAGAGTTCCATTA 606  
| | | | |  
QY 601 ACTTTGGCTCAGACCGATTAATCTTCGGCAAAAGGATGGGAGAGTTGGAAATTTGAGCTG 660  
| | | | |  
Db 607 ACTTTGGCTCAGACCGATTAATCTTCGGCAAAAGGATGGGAGAGTTGGAAATTTGAGCTG 666  
| | | | |  
QY 661 GAGACTCAGAGGAGCTTTGTTGGTCTCAAGCTGTTCCGTAATCTCAACCAAGATGCT 720  
| | | | |  
Db 667 GAGACTCAGAGGAGCTTTGTTGGTCTCAAGCTGTTCCGTAATCTCAACCAAGATGCT 726  
| | | | |  
QY 721 TTGTGACAACAACTGTGTCTGCAAGTTTTCATCCCGTGAATCCGACCCGAGCTGACCA 780  
| | | | |  
Db 727 TTGTGACAACAACTGTGTCTGCAAGTTTTCATCCCGTGAATCCGACCCGAGCTGACCA 786  
| | | | |  
QY 781 CTGTCTTAAGTCCGGAACCTTAGACAAGAACACCGTGGGCTACCTGCAGTGGCCATGGGGTA 840  
| | | | |  
Db 787 CTGTCTTAAGTCCGGAACCTTAGACAAGAACACCGTGGGCTACCTGCAGTGGCCATGGGGTA 846  
| | | | |  
QY 841 TCCAGTCAAGCATGAACATAGCATCGTCCGAGACATTAACCAAGCATTCACTGTGG 900  
| | | | |  
Db 847 TCCAGTCAAGCATGAACATAGCATCGTCCGAGACATTAACCAAGCATTCACTGTGG 906  
| | | | |  
QY 901 CCTGCAAGCTGGGAATCCCTCACTCTTTGCACTGATCAGCTATCAGACAAATTCAG 960  
| | | | |  
Db 907 CCTGCAAGCTGGGAATCCCTCACTCTTTGCACTGATCAGCTATCAGACAAATTCAG 966  
| | | | |  
QY 961 ATGACGATCAGACTCCGTTGAAGAGATCCCTCAAGAGAGCTTTTGGAGCGTGTGG 1020  
| | | | |  
Db 967 ATGACGATCAGACTCCGTTGAAGAGATCCCTCAAGAGAGCTTTTGGAGCGTGTGG 1026  
| | | | |  
QY 1021 AGTACGAGCTGAGAGAGATCTCCAGGACAGCGTTTGGTGGAGCTGTGAGCGTTG 1080  
| | | | |  
Db 1027 AGTACGAGCTGAGAGAGATCTCCAGGACAGCGTTTGGTGGAGCTGTGAGCGTTG 1086  
| | | | |  
QY 1081 GAGTTCCACAGGGGCTTTCTCTCAAAAGTCAAGCTCAACAGGGCCAGTCAACATCTCT 1140  
| | | | |  
Db 1087 GAGTTCCACAGGGGCTTTCTCTCAAAAGTCAAGCTCAACAGGGCCAGTCAACATCTCT 1146  
| | | | |  
QY 1141 TCCCTATTCACCTTGAAGGACAGCTTCTGCCAGCGCAGTGTCTATGCAACGTTGGG 1200  
| | | | |  
Db 1147 TCCCTATTCACCTTGAAGGACAGCTTCTGCCAGCGCAGTGTCTATGCAACGTTGGG 1206  
| | | | |  
QY 1201 CTCTAGTGTCTACTTTGCGATGCAACGCTGTGATCAATCAACATCACTCAAGGCTCAGA 1260  
| | | | |  
Db 1207 CTCTAGTGTCTACTTTGCGATGCAACGCTGTGATCAATCAACATCACTCAAGGCTCAGA 1266  
| | | | |  
QY 1261 AAGAGAAAGAAATTTGAGAAAGCAGAGGAAAGCCCGCCAGATGTGCTGCAGAAAGC 1320  
| | | | |

Db 1267 AAGAGAAAGAAATTTGAGAAAGCAGAGGAAAGCGCGCCACCGATGTGTCAGAAAGAAC 1326  
| | | | |  
QY 1321 AAGAGCGGAGTCCCGTGTCCGCTGATGCAAGAAATCTGTCCGAAGGAATATTGGGCGAG 1380  
| | | | |  
Db 1327 AAGAGCGGAGTCCCGTGTCCGCTGATGCAAGAAATCTGTCCGAAGGAATATTGGGCGAG 1386  
| | | | |  
QY 1381 AAGAGTCCAGAAATGGGCTCATCATGTCAATGTGCTGTGATCGGGAAGTTGTCAATGACA 1440  
| | | | |  
Db 1387 AAGAGTCCAGAAATGGGCTCATCATGTCAATGTGCTGTGATCGGGAAGTTGTCAATGACA 1446  
| | | | |  
QY 1441 AAGAGAGAAAGCGAGAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1500  
| | | | |  
Db 1447 AAGAGAGAAAGCGAGAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1506  
| | | | |  
QY 1501 TGAAGACTCCGAAGCTCATCCCTCAGAGAGCCCTCCAAAGGCTGGGCTGCTGCTTTATG 1560  
| | | | |  
Db 1507 TGAAGACTCCGAAGCTCATCCCTCAGAGAGCCCTCCAAAGGCTGGGCTGCTGCTTTATG 1566  
| | | | |  
QY 1561 ACCCGTGTGGGGGAGAGAAACCTGAAGTGTCTATCAGTTCCGGGCGTCTGCTC 1620  
| | | | |  
Db 1567 ACCCGTGTGGGGGAGAGAAACCTGAAGTGTCTATCAGTTCCGGGCGTCTGCTC 1626  
| | | | |  
QY 1621 ATCAGGTATGTGTCTGACAGTGAAGCCCTCCGATACCAAGCAAGTCCACAGATCG 1680  
| | | | |  
Db 1627 ATCAGGTATGTGTCTGACAGTGAAGCCCTCCGATACCAAGCAAGTCCACAGATCG 1686  
| | | | |  
QY 1681 ATACAGATGATTAACCTGCAAGAACCAAGATTTTAAAGGCGGCAAAATCTTTCCCT 1740  
| | | | |  
Db 1687 ATACAGATGATTAACCTGCAAGAACCAAGATTTTAAAGGCGGCAAAATCTTTCT 1746  
| | | | |  
QY 1741 GGGAGCTCAAAATTTGAAATGAAAAAACCGACATCAGATGTTTTATTTATTA 1800  
| | | | |  
Db 1747 GGGAGCTCAAAATTTGAAATGAAAAAACCGACATCAGATGTTTTATTTATTA 1806  
| | | | |  
QY 1801 TTATTATGAAGTGTGATCAATTATCAATTATGTAAGGACATGCACACCCAGCTT 1860  
| | | | |  
Db 1807 TTATTATGAAGTGTGATCAATTATCAATTATGTAAGGACATGCACACCCAGCTT 1866  
| | | | |  
QY 1861 TTGAAGGATCTGGGGTATGACATGAGAGCCCACTGGGAAACAGACTGCAGCCCTGGGC 1920  
| | | | |  
Db 1867 TTGAAGGATCTGGGGTATGACATGAGAGCCCACTGGGAAACAGACTGCAGCCCTGGGC 1926  
| | | | |  
QY 1921 CATGCTGTTTTCCCAAGGATCAGTTCTGAGAGGAAGGGCTCTGAGCCCTGACTCGCTG 1980  
| | | | |  
Db 1927 CATGCTGTTTTCCCAAGGATCAGTTCTGAGAGGAAGGGCTCTGAGCCCTGACTCGCTG 1986  
| | | | |  
QY 1981 TGTCCGAGCAACGCTGTGACCGAGCCGCGCTGTAGTTCTTGCTGGTCTGGA 2040  
| | | | |  
Db 1987 TGTCCGAGCAACGCTGTGACCGAGCCGCGCTGTAGTTCTTGCTGGTCTGGA 2046  
| | | | |  
QY 2041 GGTGCTGAGAGCAACCTGCGCTCAACCAAGAGGTGAGCACTTGCAGTCCAGCC 2100  
| | | | |  
Db 2047 GGTGCTGAGAGCAACCTGCGCTCAACCAAGAGGTGAGCACTTGCAGTCCAGCC 2106  
| | | | |  
QY 2101 TGAACATGGAAGAAACAACCTGAAGAGCAGGCTCCGCTCAGGAGCCTCTGCTGTG 2160  
| | | | |  
Db 2107 TGAACATGGAAGAAACAACCTGAAGAGCAGGCTCCGCTCAGGAGCCTCTGCTGTG 2166  
| | | | |  
QY 2161 CTGGCTTCCATGACCACTCTCTCTGTAATATTAAGTTGATGAGAGAGAT 2220  
| | | | |  
Db 2167 CTGGCTTCCATGACCACTCTCTCTGTAATATTAAGTTGATGAGAGAGAT 2226  
| | | | |  
QY 2221 GGGGGTTTAAACGCTTTTATCTGAGAACAAAGGGGTTTGAATTAAGTGTCTT 2280  
| | | | |  
Db 2227 GGGGGTTTAAACGCTTTTATCTGAGAACAAAGGGGTTTGAATTAAGTGTCTT 2286  
| | | | |  
QY 2281 TTTTCCCACTCCCAAGCTGTCTCAAGTCAATTCACCGGCGCTCGGCTTGGGACAGGT 2340  
| | | | |  
Db 2287 TTTTCCCACTCCCAAGCTGTCTCAAGTCAATTCACCGGCGCTCGGCTTGGGACAGGT 2346  
| | | | |  
QY 2341 AGTGAATCTCCGATCCAGGGCTTAGCCCTGACACAGGTGGCTTCCGATCCGAGTGG 2400  
| | | | |



```
Db 2347 AGTGTAACTCCCGATCCAGGGGCTAGCCCTGACACAGTGGCTTCCGATATCCGGTGG 2406
Qy 2401 GAAATACCCCTGCGACACAGGGGCTTGAAGTGGCTGTGTCCTCCACCCGCTGCACAC 2460
Db 2407 GAAATACCCCTGCGACACAGGGGCTTGAAGTGGCTGTGTCCTCCACCCGCTGCACAC 2466
Qy 2461 CCACCTCCAGAGTCAAGTGTGGGCAAGGGGAGCTCAAGAGAGACAGGAGGCTTGG 2520
Db 2467 CCACCTCCAGAGTCAAGTGTGGGCAAGGGGAGCTCAAGAGAGACAGGAGGCTTGG 2526
Qy 2521 CAAGACATCAGACACACCAACCCAAAGGCTGAGACCCAGGCGCCGCGCTGGTACCA 2580
Db 2527 CAAGACATCAGACACACCAACCCAAAGGCTGAGACCCAGGCGCCGCGCTGGTACCA 2586
Qy 2581 GCAGTGGACATGACAGCTCCCGCTCTCTGACAGGTTCAGGCTCTCAAGAAACACAGG 2640
Db 2587 GCAGTGGACATGACAGCTCCCGCTCTCTGACAGGTTCAGGCTCTCAAGAAACACAGG 2646
Qy 2641 CCTGTGCTCCGGAGCTTCTTCAAGACCTTCTCAAGGTGCGACCTTGGAGTCAAGAT 2700
Db 2647 CCTGTGCTCCGGAGCTTCTTCAAGACCTTCTCTCAAGGTGCGACCTTGGAGTCAAGAT 2706
Qy 2701 GCAGCGAGCTAGAGACCCCTCAAGGCTGAGACCTCGGCTGACAGTAAAGTTAGTGAAG 2760
Db 2707 GCAGCGAGCTAGAGACCCCTCAAGGCTGAGACCTCGGCTGACAGTAAAGTTAGTGAAG 2766
Qy 2761 CCTGTCTCTCGGGGCTTGGAAAGTGGCAAGCTCAAGTGTCTTGTGCTGACCTTGGAGCA 2820
Db 2767 CCTGTCTCTCGGGGCTTGGAAAGTGGCAAGCTCAAGTGTCTTGTGCTGACCTTGGAGCA 2826
Qy 2821 AGCGCGGACAGTGTGGGCTGAGACAGCTGGCGCGGGGGGCGCCAGAGCTGGCGGCT 2880
Db 2827 AGCGCGGACAGTGTGGGCTGAGACAGCTGGCGGGGGGCGCCAGAGCTGGCGGCT 2886
Qy 2881 CCAGCCCAACCCCAAGCTGTGCTGAGTCAAGGCAAGCTTCCAGCACTGTGATCTGAGT 2940
Db 2887 CCAGCCCAACCCCAAGCTGTGCTGAGTCAAGGCAAGCTTCCAGCACTGTGATCTGAGT 2946
Qy 2941 AAGCGCTAAGAACTCTCTCTCTGCTGCTTGAAGAGCAAGTGGGCTGTCCAAATCTGTA 3000
Db 2947 AAGCGCTAAGAACTCTCTCTCTGCTGCTTGAAGAGCAAGTGGGCTGTCCAAATCTGTA 3006
Qy 3001 ACATTCACTCCATTTTTTAAAGGTTCTCTGAGGCGCCCAAGCGCGGAGCGGGT 3060
Db 3007 ACATTCACTCCATTTTTT-AAAGAGTTTCTCTGAGGCGCCCAAGCGCGGAGCGGT 3065
Qy 3061 GAGCGTGTGTGATGAGCTGGGCGCCGCGGCTTCCGCTGCGCTGTGCGCAGAGTGT 3120
Db 3066 GAGCGTGTGTGATGAGCTGGGCGCCGCGGCTTCCGCTGCGCTGTGCGCAGAGTGT 3125
Qy 3121 TCTGGGCAACCATCTCTGCGTTTCATTGCAATGCAATGTAAGAGGCACTACACCA 3180
Db 3126 TCTGGGCAACCATCTCTGCGTTTCATTGCAATGCAATGTAAGAGGCACTACACCA 3185
Qy 3181 ATAAACCTTCTGAAAGCAGAAAAA 3213
Db 3186 ATAAACCTTCTGAAAGCAGAAAAA 3218
```

```
RESULT 2
US-10-104-047-1796
; Sequence 1796, Application US/10104047
; Publication No. US20030236392A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: NO. US20030236392A1el full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; PRIOR FILING DATE: 2002-03-25
; PRIORITY APPLICATION NUMBER:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
```

```
; SEQ ID NO 1796
; LENGTH: 2454
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-104-047-1796

Query Match 57.6%; Score 1851.6; DB 16; Length 2454;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1865; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

Qy 1338 GTCGGGTGATGACAGAAATGTGTCCGAAGGATTAATGAGCAGAAAGTCCAGAAATGGC 1397
Db 585 GTCGGGTGATGACAGAAATGTGTCCGAAGGATTAATGAGCAGAAAGTCCAGAAATGGC 644
Qy 1398 CTATCATGTCTCAATGCTGTGTACGGGAAGTTTGTCAATGACAAAGACAGAAAGCGAG 1457
Db 645 CTATCATGTCTCAATGCTGTGTACGGGAAGTTTGTCAATGACAAAGACAGAAAGCGAG 704
Qy 1458 AAGGTGAAGGTGATGATGACGATGTCGCTGCAAGTGCCTGGTGAAGGACTGGAAGCTC 1517
Db 705 AAGGTGAAGGTGATGATGACGATGTCGCTGCAAGTGCCTGGTGAAGGACTGGAAGCTC 764
Qy 1518 ATCTCAAGGAGGCTTCCAAAGGCTGGCTGCTGCTTTTATGACCCGCTGTGGGGAA 1577
Db 765 ATCTCAAGGAGGCTTCCAAAGGCTGGCTGCTGCTTTTATGACCCGCTGTGGGGAA 824
Qy 1578 GAGAAAGACTGAAGTGTCTTATCAAGTTCGGGGGCTCTGCATCAGGTGATGTGCTG 1637
Db 825 GAGAAAGACTGAAGTGTCTTATCAAGTTCGGGGGCTCTGCATCAGGTGATGTGCTG 884
Qy 1638 GACAGTGAAGGCTTCCGAGATACCAAGCAGTCCCAAGATGATGATGATGATTAAGT 1697
Db 885 GACAGTGAAGGCTTCCGAGATACCAAGCAGTCCCAAGATGATGATGATGATTAAGT 944
Qy 1698 GCCAAGAACAGATTTTAAAGGCGCAAAAAATCTTCTCTGGAGTCAAAATTTG 1757
Db 945 GCCAAGAACAGATTTTAAAGGCGCAAAAAATCTTCTCTGGAGTCAAAATTTG 1004
Qy 1758 GAAATGAAAAAACCCAGACATCAAGTGTTTTATTTATTTATTTATTAAGAGTGT 1817
Db 1005 GAAATGAAAAAACCCAGACATCAAGTGTTTTATTTATTTATTTATTAAGAGTGT 1064
Qy 1818 ACCATTATCAATTAATGAAAGGACATGACAGACACCCACTTTTGAAGGCTGGGGT 1877
Db 1065 ACCATTATCAATTAATGAAAGGACATGACAGACACCCACTTTTGAAGGCTGGGGT 1124
Qy 1878 AGGACTGAGGCAAGCCCACTGGGAAACCAAGACTGAGCTGAGCCCATAGGCTGTTTCCCA 1937
Db 1125 AGGACTGAGGCAAGCCCACTGGGAAACCAAGACTGAGCTGAGCCCATAGGCTGTTTCCCA 1184
Qy 1938 GATCAAGTCTTGGAGGAAAGGCTGTGAGCTGTGACTCGGCTGTGCCAGACACGTG 1997
Db 1185 GATCAAGTCTTGGAGGAAAGGCTGTGAGCTGTGACTCGGCTGTGCCAGACACGTG 1244
Qy 1998 CTGACCGCAACCCCGCGCTGTGATGTTCTTGGCTGGGCTCTGAGAGTGTCTGTGAGAC 2057
Db 1245 CTGACCGCAACCCCGCGCTGTGATGTTCTTGGCTGGGCTCTGAGAGTGTCTGTGAGAC 1304
Qy 2058 CTGCGCTCAGCAGAGAGCTGAGCACTTCTGAGCTCAGCTGAAACATGAGGAAAC 2117
Db 1305 CTGCGCTCAGCAGAGAGCTCAGCTGAGCACTTCTGAGCTCAGCTGAAACATGAGGAAAC 1364
Qy 2118 CTGAAAAGCAGGAGGCTTCCGCTGAGGAGGCTGTGCTGTGCTGCTTCCATGACA 2177
Db 1365 CTGAAAAGCAGGAGGCTTCCGCTGAGGAGGCTGTGCTGTGCTGCTTCCATGACA 1424
Qy 2178 CTTCTCTCTGTAATTAATTAATGCTGTAATCTGAGAGAGATGCGGGTTTATTAAC 2237
Db 1425 CTTCTCTCTGTAATTAATTAATGCTGTAATCTGAGAGAGATGCGGGTTTATTAAC 1484
Qy 2238 CTTTATCTGAGAAACAAACGGGTTTGAATTAATGCTGCTTCTTCTCCCACTCCAGAG 2297
Db 1485 CTTTATCTGAGAAACAAACGGGTTTGAATTAATGCTGCTTCTTCTCCCACTCCAGAG 1544
```

QY 2298 CTGCTCAAGTCATTTCCACGGCCCCCTCGGCTTGGGACAGAGTATGTAATCCCGATCC 2357  
| | | | |  
Db 1545 CTGCTCAAGTCATTTCCACGGCCCCCTCGGCTTGGGACAGAGTATGTAATCCCGATCC 1604  
| | | | |  
QY 2358 CAGGGGCTAGCCCTGACACAGGTGGCTTCCGATATCCCGGTGGGAAAAACGCTGCAC 2417  
| | | | |  
Db 1605 CAGGGGCTAGCCCTGACACAGGTGGCTTCCGATATCCCGGTGGGAAAAACGCTGCAC 1664  
| | | | |  
QY 2418 AGCGGGGTTAGCTGGGCTGTGTCCCTCCACCGGCTTGACACACCTTCAAGATGAC 2477  
| | | | |  
Db 1665 AGCGGGGTTAGCTGGGCTGTGTCCCTCCACCGGCTTGACACACCTTCAAGATGAC 1724  
| | | | |  
QY 2478 TGGTGGGCAAGGAGGAGCTCAAGAGACAGAGAGCGGCTTGGCAAGACATCAGACAC 2537  
| | | | |  
Db 1725 TGGTGGGCAAGGAGGAGCTCAAGAGACAGAGAGCGGCTTGGCAAGACATCAGACAC 1784  
| | | | |  
QY 2538 CCAACCCAAAGGCTGAGACCCAGGCGCGGCTGGTACCCAGCAGGTGGCACTGCAG 2597  
| | | | |  
Db 1785 CCAACCCAAAGGCTGAGACCCAGGCGCGGCTGGTACCCAGCAGGTGGCACTGCAG 1844  
| | | | |  
QY 2598 TCCCGGCTCTGACAGGTCCAGGCTCTTCAAGAGAACCCAGGCTGTGTCCGAGGCT 2657  
| | | | |  
Db 1845 TCCCGGCTCTGACAGGTCCAGGCTCTTCAAGAGAACCCAGGCTGTGTCCGAGGCT 1904  
| | | | |  
QY 2658 TCCCTTCAAGACCTTCCCTCCAGTGGCCACTTGGGATGACAGATGACGCGGAGCTAG 2717  
| | | | |  
Db 1905 TCCCTTCAAGACCTTCCCTCCAGTGGCCACTTGGGATGACAGATGACGCGGAGCT 1964  
| | | | |  
QY 2718 CCCTTCAAGGCTGAGACCTTGGCTGACAGTAAAGTTAGTGAAGGCTGTCTCTGGG 2777  
| | | | |  
Db 1965 CCCTTCAAGGCTGAGACCTTGGCTGACAGTAAAGTTAGTGAAGGCTGTCTCTGGG 2824  
| | | | |  
QY 2778 GGAAGTGGGAGCCATGATGCTTGTCTTGTCTTGAAGGAGGAGGCGGAGAGGTGT 2837  
| | | | |  
Db 2025 GGAAGTGGGAGCCATGATGCTTGTCTTGTCTTGAAGGAGGAGGCGGAGAGGTGT 2084  
| | | | |  
QY 2838 GGGTGAAGACAGTGGGCGGCGGAGGCGGAGGCTTCCAGGCGGAGGCGGAGGCT 2897  
| | | | |  
Db 2085 GGGTGAAGACAGTGGGCGGCGGAGGCGGAGGCTTCCAGGCGGAGGCGGAGGCT 2144  
| | | | |  
QY 2898 GTTGTGAAGTCAAGGCGGAGGCGGAGGCTTCCAGGAGTGTGAAGTGAAGGCTTCC 2957  
| | | | |  
Db 2145 GTTGTGAAGTCAAGGCGGAGGCGGAGGCTTCCAGGAGTGTGAAGTGAAGGCTTCC 2204  
| | | | |  
QY 2958 TTTCTCTGCTTTGAAAAAGAGTGGGTTGTCCAAATTTCTGAATCATTTCTCCATTT 3017  
| | | | |  
Db 2205 TTTCTCTGCTTTGAAAAAGAGTGGGTTGTCCAAATTTCTGAATCATTTCTCCATTT 2264  
| | | | |  
QY 3018 TTTAAAAAGGTTTCTGACGCGGCGGAGGCGGAGGCGGAGGCTTCCAGGAGTGTG 3077  
| | | | |  
Db 2265 TTTAAAAAGGTTTCTGACGCGGCGGAGGCGGAGGCGGAGGCTTCCAGGAGTGTG 2223  
| | | | |  
QY 3078 AGCTTGGGCGGCGGAGGCTTCCGCTGCGGCTTGTGCGAGGTGTCTTGGGAGCCATCT 3137  
| | | | |  
Db 2324 AGCTTGGGCGGCGGAGGCTTCCGCTGCGGCTTGTGCGAGGTGTCTTGGGAGCCATCT 2283  
| | | | |  
QY 3138 TGGGTTTCAATTTGACAGTCTGTACAGAGGAGCATCACAATAATCTTCTGAAA 3197  
| | | | |  
Db 2384 TGGGTTTCAATTTGACAGTCTGTACAGAGGAGCATCACAATAATCTTCTGAAA 2443  
| | | | |  
QY 3198 GCAGAAAAAA 3207  
| | | | |  
Db 2444 GCAGAAAAAA 2453  
| | | | |

RESULT 3  
US-10-062-674-1802/c

; Sequence 1802, Application US/10062674

; Publication No. US20040005559A1

; GENERAL INFORMATION:

; APPLICANT: Loring, Jeanne F.; Kaser, Matthew R.

; TITLE OF INVENTION: MARKERS OF NEURONAL DIFFERENTIATION AND MORPHOGENESIS

; FILE REFERENCE: PA-0026-1 CIP  
; CURRENT APPLICATION NUMBER: US/10/062,674  
; CURRENT FILING DATE: 2002-01-30  
; PRIOR APPLICATION NUMBER: US 09/625,102  
; PRIOR FILING DATE: 2000-07-24  
; NUMBER OF SEQ ID NOS: 2217  
; SOFTWARE: PERL Program  
; SEQ ID NO 1802  
; LENGTH: 3600  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; NAME/KEY: misc.feature  
; OTHER INFORMATION: Incyte ID No. US20040005559A1 253870.6c  
; NAME/KEY: unsure  
; LOCATION: (1) ... (3600)  
; OTHER INFORMATION: a, t, c, g, or other  
US-10-062-674-1802

Query Match 50.1%; Score 1608.8; DB 16; Length 3600;

Best Local Similarity 83.1%; Pred. No. 0;

Matches 2839; Conservative 0; Mismatches 363; Indels 213; Gaps 80;

QY 1 GAAAGTTGCGAAGATGCGGAGCGGCTTGAAGGAGGAGGAGCTGACATGAAGACTATT 60  
| | | | |  
Db 3594 GAAAGTTGCGAAGATGCGGAGCGGCTTGAAGGAGGAGGAGCTGACATGAAGACTATT 3535  
| | | | |  
QY 61 ACTGTTGCTGAACGTCGCGAGGAGGCTC-TTCTGAAGACTGAAGC-TGCCACCG 118  
| | | | |  
Db 3534 ACTGTTGCTGAACGTCGCGAGGAGGCTCTTCTGAAGACTGAAGAACTTGCTACCG 3475  
| | | | |  
QY 119 GA--GGCTGTATGCTTAC--ATCCAGACAGACAGAG--CCAGAGCTCAAGTAC 174  
| | | | |  
Db 3474 GAAGCTGTATGCTTACCAATCCAGACAGACAGAGTCCAGAGCTCAAGTAC 3415  
| | | | |  
QY 175 AGGCGGAACGATGTTTAACTTGTTCACAG--GCTTATGAAGTGTATGACCCCA 233  
| | | | |  
Db 3414 AGGCGGAACGATGTTTAACTTGTTCACAGAGCTTATGAAGTGTATGACCCCA 3355  
| | | | |  
QY 234 ACCAGG--CCATCTATGATATATATGGAAGAGAGCTGAAATGAAAGATGGAGGT 292  
| | | | |  
Db 3354 ACCAGGCGCATCTATGATATATATGGAAGAGAGCTGAAATGAAAGATGGAGGT 3295  
| | | | |  
QY 293 TGTGAAGAGAGAGAACCCCTGC--TGAATTCGAGAGAGTTGAGCGGCTG---CAGA 348  
| | | | |  
Db 3294 TGTGAAGAGAGAGAACCCCTGCCTTGAATTCGAGAGAGTTGAGCGGCTGAGAGAG 3235  
| | | | |  
QY 349 GAGAGAGAGAGAGAGATGACAGAGCAATCCAAAGGAAAGATCAGCGTTG 408  
| | | | |  
Db 3234 GAGAGAGAGAGAGAGATGACAGAGCAATCCAAAGGAAAGATCAGCGTTG 3175  
| | | | |  
QY 409 GAGTAGAT--CCACCGACTTTTGAATGCTATGATGAGAGATGAAAGATGTTC--GG 466  
| | | | |  
Db 3174 GAGTAGATCCACCGACTTTTGAATGCTATGATGAGAGATGAAAGATGTTC--GG 3115  
| | | | |  
QY 467 CAGTAGCTTT--CCGAGATGAAATTAATTAATGAACAATCCAGATCCATT--GAGGCA 524  
| | | | |  
Db 3114 CAGTAGCTTTCCCGAGATGAAATTAATTAATGAACAATCCAGATCCATTGAGGCA 3055  
| | | | |  
QY 525 CCCTTGACAGACAGACAGACATCTCTCTGGAAGGCTTCAACCCAGAAATGA---581  
| | | | |  
Db 3054 CCCTTGACAGACAGACAGACATCTCTCTGGAAGGCTTCAACCCAGAAATGA---581  
| | | | |  
QY 582 ---AATGAGAGGTTCCATTACTT---CGGCTCAGACAGATTAATCTTGGCAAGG 634  
| | | | |  
Db 2994 GACAAATGAGAGGTTTCCATTACTTGGCGCTCCAGACAGATTAATCTTGGCAAGG 2935  
| | | | |  
QY 635 ATGGGAGAGTTGGAATTTGAGCTGA-----GACCTACAGGGGCTTTGTTGG 685  
| | | | |  
Db 2934 ATGGGAGAGCTTGGACATTTGGAGCGCTGAGAACTTACAGGGGCTTTGTTGG 2875  
| | | | |  
QY 686 TCTCAAGCTGTCCGTAATCTCACACCAAGATGCTTGTGACAAACACTGTGCTGCA 745  
| | | | |

```
Db 2874 TCTCAGCTGTTCCGTAATCTCAACCAAGATGCTTTGTGACAAACAACTGTGCTGCA 2815
Qy 746 GTTT-TGATCCCGTGGAAATCC-GACCCGGCTGACCACTGTCTAGCTGGAACTTGAC 803
Db 2814 GTTTATCATCCCGTGGAAATCTGACCCCGCTGACCACTGTCTAGCTGGAACTTGAC 2755
Qy 804 AAGAACACCGTGGG-----CTACCTGCAGTGGGATGGGGATTCACGTACGCAATGAC 857
Db 2754 AAGAACACCTGTGGGCTACCTGCAGTGGGATGGGGATTCACGTACGCAATGAC 2695
Qy 858 ACTAGCATCTGTC-GGAGACTTAAACACGACACTTCACTGTGGCCCTGACAGCTGGAAAT 916
Db 2694 ACTAGCATCTGTGTAAGGCACTAGAACAGGCACTTCACTGTGGCCCTGACAGCTGGAAAT 2635
Qy 917 CCTCTACTCTTTTGCACTGATGAGTATCAGACAAATTCAGATGACATGACATGCTG 976
Db 2634 CCTCTACTCTTTTGCACTGATGAGTATCAGACAAATTCAGATGACATGACATGCTG 2575
Qy 977 TGTAAAGGATCCCTCAAGAGGCTTCTTGGACGGTGGTGGATGAGTACGGAGCTGAGAG 1036
Db 2574 TGTAAAGGATCCCTCAAGAGGCTTCTTGGACGGTGGTGGATGAGTACGGAGCTGAGAG 2515
Qy 1037 GAAATCTCCAGGACAGCGCTTTTGGTGGAGCTGTGACGGTGGAGTTCCACAGGCGCT 1096
Db 2514 AGAATCTCCAGGACAGCGCTTTTGGTGGAGCTGTGACGGTGGAGTTCCACAGGCGCT 2455
Qy 1097 TTCTCTCAATGACTCAAGCTCAACAGGGCCAGTACATATCTTCCCTATTCACCTTAC 1156
Db 2454 TTCTCTCAATGACTCAAGCTCAACAGGGCCAGTACATATCTTCCCTATTCACCTTAC 2395
Qy 1157 -GGACAGGCTTC-TGCCAGGCGCATGTTCTATGCAACCGTGGGCGCTTATGCTGCTAC 1214
Db 2394 TGGACAGGCTTCCTTGGCCAGGCGCATGTTCTATGCAACCGTGGGCGCTTATGCTGCTAC 2335
Qy 1215 TTTGCAATGACCGTTC-TGATCATCAACCATACCTCAGGCGCTCAGAAAGAGAAAT 1273
Db 2334 TTTGCAATGACCGTTCCTTGGATCATCAACCATACCTCAGGCGCTCAGAAAGAGAAAT 2275
Qy 1274 GGAGAGCAGAGGAGAAAGCGCGCCACCGATGTGCTGCAAGAAAGCAA-----GAGGC 1327
Db 2274 GGAGAGCAGAGGAGAAAGCGCGCCACCGATGTGCTGCAAGAAAGCAA-----GAGGC 2215
Qy 1328 GGAATCCGCTGTCCGGCTGATGCAAGATCTGTCCGAAGATTAATTGAGGAGAGAGTC 1387
Db 2214 GGAATCCGCTGTCCGGCTGATGCAAGATCTGTCCGAAGATTAATTGAGGAGAGAGTC 2155
Qy 1388 CAGAAATGGGCTCTCATCTGCTCAATGCTGTGAC-GGGAAATTTGTCAATGACAAAGACA 1446
Db 2154 CAGAAATGGGCTCTCATCTGCTCAATGCTGTGACGGGAAATTTGTCAATGACAAAGACA 2095
Qy 1447 GGAAGAGCAGAGAGGTGAAGGTGATTG--ACGTGACTGTGCCCCCTGACAGTGCCTGTGTA 1503
Db 2094 GGAAGAGCAGAGAGGTGAAGGTGATTGCACTGTGACCTGTGACCTGTGAGTGCCTGTGTA 2035
Qy 1504 AGGATCGAAGCTCATCTCTCAACGAGG--CTTCCGAAGGTGGGCTGCTGCTTTTATGA- 1561
Db 2034 GGGATCGAAGCTCATCTCTCAACGAGGCTTCCGAAGGTGGGCTGCTGCTTTTATGAC 1975
Qy 1562 CCCGTGTGTGGGGAGAGAAAGAACTGAAAGTCTCTATGAG--TTCGGGGGCTGCTGTC 1620
Db 1974 CCCGTGTGTGGGGAGAGAAAGAAAGCTGAAAGTCTCTATGAGTTTCCGGGGGCTGCTGTC 1915
Qy 1621 ATCAGGTGATGG-TGCTGAGACATGAGGCGCTCCGATACCAAGACAGT-----CCCA 1672
Db 1914 ATCAGGTGATGGCTGTGAGACATGAGGCGCTTCACTGAGATACATATGAGAGTCCAT 1855
Qy 1673 CAGAGATGATACAGATGATGATAA--CTGCCAAGAACAGATTTT--AAAAGCCGCAAAA 1729
Db 1854 CAGAGATGATACAGATGATGATAA--CTGCCAAGAACAGATTTT--AAAAGCCGCAAAA 1795
Qy 1730 AATC--TTTTCTGGAGTCTACAAAT-TTGAATGAAAAAACCCAGACATCATGATTT 1786
Db 1730 AATC--TTTTCTGGAGTCTACAAAT-TTGAATGAAAAAACCCAGACATCATGATTT 1786
Db 1794 AATCTTTGATGAGAGTCTACAAATGTTGAAATGAAAAAACCCAGACATCATGATTT 1725
Qy 1787 TTTATTTTATTTATTTATTTATAGAAAGTGGTACATTTATCA-ATTATGTG-AAGGACAT 1844
Db 1734 TTTATTTTATTTATTTATTTATAGAAAGTGGTACATTTATGATATGATAAAGGACAT 1675
Qy 1845 GCAGACACCCGAGTTTGAAGGGG--CTGGGGGGTGAAGTGAAGGACCCCACTGGGAAAC 1903
Db 1674 GCAGACACCCGAGTTTGAAGGGGCTGGGGGGTGAAGTGAAGGCTCCCTTACGGGGTG 1615
Qy 1904 CAGACTCAGCTGGC-----CAATGCTGTTTTTCCAGAGATGAG--TTCCTGG 1951
Db 1614 AACAGACTTGAAGGCGTGGTGCATTTGGGTGTGTGTGCCAAGAGTGTAGCTGCGGG 1555
Qy 1952 AAGGAAAGGCTCTGGGCC--TGACTCCGTGTGTCCGAGCACATGCTGACCGGAGC 2008
Db 1554 CAGGAAAGGCTCTGGAGCCCTGACATCCGCTGTGTCCGAGCACATGCTGACCGGAG 1495
Qy 2009 CCGCGCGCCCTGT--AGTCTGGCTGGGCT--GAGAGTGTGTGAGAGACCTGCGCT 2064
Db 1494 CCGCGCGCCCTGTAGTGTCTGGCTGGGCTGTAGAGAGTGTGTGAGAGACCTGCGCT 1435
Qy 2065 CACCAACAGG-AGGCTGAGCACTTCTGC--AGTCCAGCTGAACAT-GGGAAACAACTG 2120
Db 1434 CACCAACAGGAGGTGAGCACCTTCTGCAAGTCTCAAGCTGAGATGAGGAAACAACTG 1375
Qy 2121 AAAAGCA--GGACAGCTCCCGGTGAGGAG--CTGTGTGTGCT-GGCTTCCATGAC 2176
Db 1374 AAAAGCATGACAGGCTCCCGGTGAGGAGCCTGTGTGTGCTGAGGCTTCCATGAC 1315
Qy 2177 ACCCTCC-TCCGTGCAAAATTTACTGCTTAA--TCTGAGAGAGATTTGGGGTTTAA 2232
Db 1314 ACCCTCTTCTGTGCAAAATTTACTGCTTAAATGCTGTGAGAGAGATTTGGGGTTTAA 1255
Qy 2233 AACTGCTTTTATCTGAAACAAACG-----GGTTTGA 2267
Db 1254 GAAATGCTTGTGTGAGTCTGAGAGAACCAACACCGGGGNTTTTGAAGAGTTAG 1195
Qy 2268 ATTATGCTCTTTTTCCTCCCATCTCCA-GAGTGTCAAGTCAATTCACCG--CCCC 2323
Db 1194 GTCGCTTGTGTGTGCTCCCATGCTCCCATGAGTGTCAATTCACACAGGACCCCT 1135
Qy 2324 TCGGCTTTGGGACAGGGATGATTAATCCG--ATCCAGGGCTTACCCCTGACACAGG 2379
Db 1134 CGGCTTAAAGGACAGGGATGATTAATCTGCTGATCTCCACAGGGCTTAACTGAAACAA 1075
Qy 2380 TGGCTTCC-----GTATCCCGGTGGAGAAACGCTGACAC--AGCGGGCTTGA 2428
Db 1074 TGGTGGGCTTCCCGGTAATCCCGGTGGGAAACGCTTGTGACACAGCGGGCTTGA 1015
Qy 2429 GCT-GGCTGTGTCTT-CCACCGCTGACCAACCACTTCA--GAGTCAAGTGTGG 2484
Db 1014 GCTGGGCTGTGTCTTCAACCGCTGACCAACCACTTCAAGAGGTCAAGTGTGG 955
Qy 2485 C--AAGGAGAGCTCAAGAG--ACAGACAGAGGCTTGGAAAGCA--TCAACA 2534
Db 954 CAAAGGAGAGCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 895
Qy 2535 CACCAACCCAAAGAGCTGAGACCCA-GGCCGGGCGTGTGTATCCAGAGAGTGTGACTG 2593
Db 894 CACCAACCCAAAGAGGTGAGACCCAAGGCGCGGCTGTGTATCCAGAGAGTGTGACTG 835
Qy 2594 CAGTCCCGCTCTCTGAGG-TCAAGCTCTTCAAGAGAACACAGAGGCTGTGTCTCG 2652
Db 834 CAGTCCCGCTCTCTGAGGTTCAAGGCTCTTCAAGAGAACACAGAGGCTGTGTCTCG 775
Qy 2653 AGCTTCTCTTCAAGACCTTCTCAAGTGTGCACTTGGATGAGATGAGAGGAGCTA 2712
Db 774 AGCTTCTCTTCAAGACCTTCTCAAGTGTGCACTTGGATGAGATGAGAGGAGCTA 715
Qy 2713 GG--ACCCCTCAAGGCTGAGACTCGGCTGAGTAAAGTTACGTAGAGGCTGTCTCT- 2769
Db 714 GGAACCCCTCAAGGCTGAGACTCGGCTGAGTAAAGTTACGTAGAGGCTGTCTCTA 655
```

OY	2770	CGGGGCGCTGGAAGTGGCAGCGCATCATGTTGCTCTTGCTGACGCC--TGGAGACAAAGCGCG	2827
Db	654	CGGGGCGCTGGAAGTGGCAGCAACCATCATGTTGCTCTTACTGACCCCTTGCGAACAAGCGCG	595
OY	2828	CACAGTGTGTGCTGAGACA-----GCTGGCGCGGGGGGCCCCAA-----GCTGC	2872
Db	594	CACAGTGTGTGGCTCTGAGACACAGCTGGGCCCCCGGGGGGGGCCCCAAAGCTGACGC	535
OY	2873	GCCGGGCGCTCCAGGCCACCCACAGACTGTT-GCTGAAGTACAGGCCCAACCT-CCCGAGCACTG	2930
Db	534	CGGGCACTCCAGGCCACCCACAGCTGTGTAGCTGAAGTACAGGCTCTCCCTCCCGCAGAACTG	475
OY	2931	GTAT-CTGAGTAAACGGCTTAAGAACCTCTCTCTCT-----CTGGTTTGGAAAAGCAGTTTC--	2982
Db	474	GTATACTGATTAACCGGCTTAAGGGAACCGTCTCTCTCTCTGGGTTTGGAAAAGCAGTTTCAG	415
OY	2993	GGGTTGTCCAAATCTGTAAACATTCATCTCCACTTTTAAAAAAGTTTCTGTG-----AC	3037
Db	414	GGTGTGTCCCAANTCTGTAAACATTCATCTCCACTTTTAAAAAAGTTTCTGTGACAG	355
OY	3038	GGCCCAACGGCCCGAGCGCG--GGTGAAGCTGCGTTGCATGAGCTGGGCCCGCGAGCTT	3095
Db	354	GCCCCAGGAACTCCAGCGCCCGGGGTGAGGTCCGTTGTGATGAGCTGGGCTGACAGGCTAT	295
OY	3096	CCC--GTGGGCTCTGTGCGAGAGTGTCTTGTGG--CACCCAATCTCTGCGTTTCATT	3149
Db	294	CCCGGTGGGCACTGTGACCAAGGTGTCTTGTGGGACCCCAATCTGTGCGTTTCATT	235
OY	3150	GCAGTGCAGCTG--TACAGAAAGGCACTGACCAATTAACCTTTCTGAAAGCAGAA	3202
Db	234	GCAGTGCAGCTGTACAGAAAGGCACTGACCAATTAACCTTTCTGTAAGCAGAA	180

```

RESULT 4
US-10-264-049-601
; Sequence 601, Application US/10264049
; Publication No. US20040005579A1
; GENERAL INFORMATION:
; APPLICANT: Birse et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA133p1
; CURRENT APPLICATION NUMBER: US/10/264,049
; CURRENT FILING DATE: 2002-10-04
; PRIOR APPLICATION NUMBER: PCT/US01/18569
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: US 60/209,467
; PRIOR FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 4360
; SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 601
; LENGTH: 1454
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-264-049-601

```

Query Match	44.0%	Score 1414.6;	DB 16;	Length 1454;
Best Local Similarity	99.6%	Pred. No. 0;		
Matches 1425; Conservative	3;	Mismatches	2;	Indels 1; Gaps 1;

[illegible]

Db	184	CTGGCCCTGACTCCGCTGTGTCTCCGAGACACAGTGTGACCCGAGCCCGCCCTGTAG	243
Oy	2023	TTCTTGGCTGGGTCTGAGGGTGTCTGTGGAGACACCTTGCCCTTACACAGAGCGTAGC	2082
Db	244	TTCTTGGCTGGGTCTGAGGGTGTCTGTGGAGACACCTTGCCCTTACACAGAGCGTAGC	303
Oy	2083	CACCTTCGAGCTCCAGCCTGAACAATGGGAAACAACTGGAAAGAGAGGACAGGCTCCGGT	2142
Db	304	CACCTTCGAGCTCCAGCCTGAACAATGGGAAACAACTGGAAAGAGAGGACAGGCTCCGGT	363
Oy	2143	CAGGAGCCCTCTGTGTGTGTGGCTTCCATGACACACTCTCTCTGTGAATATTAATGCG	2202
Db	364	CAGGAGCCCTCTGTGTGTGTGGCTTCCATGACACACTCTCTCTGTGAATATTAATGCG	423
Oy	2203	TTGAATCTGAGACAGATTCGGGGTTTAAATACTGCTTTTATCTGAAACAAACGGGTT	2262
Db	424	TTGAATCTGAGACAGATTCGGGGTTTAAATACTGCTTTTATCTGAAACAAACGGGTT	483
Oy	2263	TGGAATTAATGTCTGCTTTTTTCCCACTCCCAAGCTGTCTCAAGTCAATTCACCGGGCCC	2322
Db	484	TGGAATTAATGTCTGCTTTTTTCCCACTCCCAAGCTGTCTCAAGTCAATTCACCGGGCCC	543
Oy	2323	CTCGGCTTGGGACAGGGTAGTGTAACTCCCGGATCCAGGGGCTTAGCCCTGACACAGGTGG	2382
Db	544	CTCGGCTTGGGACAGGGTAGTGTAACTCCCGGATCCAGGGGCTTAGCCCTGACACAGGTGG	603
Oy	2383	CTTCCCGTATCCCGGTGGGAAAACGCCCTGCACACAGCGGGCTTGAAGCTGAGCTGTGTCC	2442
Db	604	CTTCCCGTATCCCGGTGGGAAAACGCCCTGCACACAGCGGGCTTGAAGCTGAGCTGTGTCC	663
Oy	2443	CTCCACCGCTTCGACCAACCACTCTCAGAGTGCAGTGTGTGGCAAGGGCAGCTCAGAGG	2502
Db	664	CTCCACCGCTTCGACCAACCACTCTCAGAGTGCAGTGTGTGGCAAGGGCAGCTCMAAGR	723
Oy	2503	ACAGAGCACAGGCGCTTGGCAAGACATAGACACACCCAAACCCAAAGCGTGGACCCAGG	2562
Db	724	ACAGAGCACAGGCGCTTGGCAAGACATAGACACACCCAAACCCAAAGCGTGGACCCAGG	783
Oy	2563	CCCGGCCCGTGTGATCCAGAGAGGTGGCACTGACGCTCCCGCTCTTGCAAGTCCAGCGTC	2622
Db	784	CCCGGCCCGTGTGATCCAGAGAGGTGGCACTGACGCTCCCGCTCTTGCAAGTCCAGCGTC	843
Oy	2623	CTCACAGGAAACCAAGGGCTGTGTCTCCGAGGCTTCTTCAAGACCTTCTCTCAAGTGC	2682
Db	844	CTCACAGGAAACCAAGGGCTGTGTCTCCGAGGCTTCTTCAAGACCTTCTCTCAAGTGC	903
Oy	2683	CCAATTGGGATGACGATGAGGAGGAGACTAGAACCCCTCCAGGCGCTGACACCTCGGCTG	2742
Db	904	CCAATTGGGATGACGATGAGGAGGAGACTAGAACCCCTCCAGGCGCTGACACCTCGGCTG	963
Oy	2743	CAGTAAAGTTAACGTGAGGCTGTCTCTCGGGGGCTGGAAATGGCAGCATCAGTTGTCT	2802
Db	964	CAGTAAAGTTAACGTGAGGCTGTCTCTCGGGGGCTGGAAATGGCAGCATCAGTTGTCT	1023
Oy	2803	TGCTGACCCCTTCGAGAGCAAGCCGCGCAACAGTGTGTGTGAGACACGCTGCGCGGGGCG	2862
Db	1024	TGCTGACCCCTTCGAGAGCAAGCCGCGCAACAGTGTGTGTGAGACACGCTGCGGGGGCG	1083
Oy	2863	CCCAAGTGTGGCGGGCTTCCAGGCCCAACCAAGCTGTGTGTGAAAGTAAAGCCAACTGCC	2922
Db	1084	CCCAAGTGTGGCGGGCTTCCAGGCCCAACCAAGCTGTGTGTGAAAGTAAAGCCAACTGCC	1143
Oy	2923	CAGACACTGTATCTGAGTAAACGGCTAAGAACTCTTCTCTGTGGTTTGAAGACAGTTC	2982
Db	1144	CAGACACTGTATCTGAGTAAACGGCTAAGAACTCTTCTCTGTGGTTTGAAGACAGTTC	1203
Oy	2983	GGGTTGTCCAATTCTGTAAACATTCATCCATTTTAAAGAAAGTTTCTCTGACGGCCC	3042
Db	1204	GGGTTGTCCAATTCTGTAAACATTCATTCATTTTAAAGAAAGTTTCTCTGACGGCCC	1262
Oy	3043	CAGGCGCCGAGCCGGGTGAGCGTCTGTGTGATGAGCTTGGGCCCGCGGCTTCCGTGC	3102

Db	Qy	Db	Qy
1263	CACGSCCCGAGCCGGGAGACGTCGTGTTCGATAGAGCTGGGCCCGGGCTTCCCGCG	13222	
3103	GGCCTCTGGCGGAGAGTGTCTTCTGGGGACCCATCTCTGGCTTCAATTGCGAGTCGATGTA	3162	
1323	GGCTCTGGCGGAGAGTGTCTTCTGGGGACCCATCTCTGGCTTCAATTGCGAGTCGATGTA	1382	
3163	CAGAAAGCACTCACCACATATAAACCCTTCTCTGAAAGCAGAAAAA	3213	
1383	CAGAAAGCACTCACCACATATAAACCCTTCTCTGAAAGCAGAAAAA	1433	

RESULT 5  
US-09-92

```

Sequence 487, Application/US/09925300
Patent No. US20020151681A1
GENERAL INFORMATION:
APPLICANT: Craig Rosen,
APPLICANT: Steve Ruben,
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PAL01
CURRENT APPLICATION NUMBER: US/09/925,300
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05988
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 1890
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 487
LENGTH: 1451
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (1256)
OTHER INFORMATION: n equals a,t,g, or c
US-09-925-300-487

```

Query Match	44.0%	Score 1413.6	DB 9	Length 1451
Best Local Similarity	99.5%	Pred. No. 0		
Matches 1424	Conservative	3	Mismatches	3
			Indels	1
			Gaps	1

QY	1783	IGTTTTTATTTTATATATATATATATAGAAAGGTGGTACCATTAATCAATTAATGTGAAGGAC	1842
Db	1	TGTTTTTATTTTATATATATATATATATAGAAAGGTGGTACCATTAATCAATTAATGTGAAGGAC	60
QY	1843	ATGCAGACACCCCAAGCTTTTATAGAGGTGTGTGGGGGTATGAGATGAGGACAGCCCACTGGGAA	1902
Db	61	ATGCAGACACCCCAAGCTTTTATAGAGGTGTGTGGGGGTATGAGATGAGGACAGCCCACTGGGAA	120
QY	1903	CCAGACTGACAGCTTGAGCCATAGGCTGTTTCCCAAGATCAATTCCTTGAAGGAAAGGCT	1962
Db	121	CCAGACTGACAGCTTGAGCCATAGGCTGTTTCCCAAGATCAATTCCTTGAAGGAAAGGCT	180
QY	1963	CTGGGCGCTGATCTCGCGTGTGTCCGAGAGACAGTGTGACCGGAGCGCCGCGCCTGTAG	2022
Db	181	CTGGGCGCTGATCTCGCGTGTGTGTCCGAGACAGTGTGACCGGAGCGCCGCGCCTGTAG	240
QY	2023	TTCTTGGCTGTGGGTCTGAGAGTGTCTGTGAGACACCTTCCTCAACAAGAGCGTAGC	2082
Db	241	TTCTTGGCTGTGGGTCTGAGAGTGTGTGTGAGACACCTTCCTCAACAAGAGCGTAGC	300
QY	2083	CACCTTCTGCAATCCACGCTGAAACATGGGAAACAATGAAAAGCAGCGTCTCCGGT	2142
Db	301	CACCTTCTGCAATCCACGCTGAAACATGGGAAACAATGAAAAGCAGCGTCTCCGGT	360
QY	2143	CAGGAGCGCTGTGCTGTGCTGGCTTCCCATGACACACTCTCTGCTGAATATATTAATGCG	2202
Db	361	CAGGAGCGCTGTGCTGTGCTGGCTTCCCATGACACACTCTCTGCTGAATATATTAATGCG	420
QY	2203	TTGAATCTGAGACGATTTGGCGGTTTATATAAACAATGCTTTTATCTGAGAACAAACGGTT	2262
Db	421	TTGAATCTGAGACGATTTGGCGGTTTATATAAACAATGCTTTTATCTGAGAACAAACGGTT	480

QY	2263	GGAAATTAGTCGCTTTTTCCTCCACATCCCAAGCTGTCTCAAGCATTTCCACGGCCCC	2322
Db	481	TGAAATTAATGTCCTTTTTCCTCCACATCCCAAGCTGTCTCAAGCATTTCCACGGCCCC	540
QY	2323	CTCGGCTTGGGACAGAGGTAGTGAATCTCCGATCCAGGGGCTTAGCCCTGACACAGGTG	2382
Db	541	CTCGGCTTGGGACAGAGGTAGTGAATCTCCGATCCAGGGGCTTAGCCCTGACACAGGTG	600
QY	2383	CTTCCCGTATCCCGGTGGAAAAAGCCCTGCAACAGGGGCTTAGCTGCTGTGTCC	2442
Db	601	CTTCCCGTATCCCGGTGGAAAAAGCCCTGCAACAGGGGCTTAGCTGCTGTGTCC	660
QY	2443	CTCCACGGCCTTGACACACCCCACTTCCAGGTGACGTGTGGGCAAGGGCACTCAAGG	2502
Db	661	CTCCACGGCCTTGACACACCCCACTTCCAGGTGACGTGTGGGCAAGGGCACTCAAGG	720
QY	2503	ACAGGACCAAGGCGCTTGGCAAGCATGACACACCAACCAAGGCGTGGACCCAGG	2562
Db	721	ACAGGACCAAGGCGCTTGGCAAGCATGACACACCAACCAAGGCGTGGACCCAGG	780
QY	2563	CCCGGCCGTGGTACCCAGCAGGTGGCATGTGACGTCCCGCTCTTGAGGTCCAGGCTC	2622
Db	781	CCCGGCCGTGGTACCCAGCAGGTGGCATGTGACGTCCCGCTCTTGAGGTCCAGGCTC	840
QY	2623	CTCACAGGAACAACAAGGCGCTGTGCTCCGAGCCTTCTTCAGACCTTCTCTCAAGTGC	2682
Db	841	CTCACAGGAACAACAAGGCGCTGTGCTCCGAGCCTTCTTCAGACCTTCTCTCAAGTGC	900
QY	2683	CCACTTGGGATGACGAATGACAGCGGAGCTAGAGACCCCTCCACGGCCTGACCTCGGCTG	2742
Db	901	CCACTTGGGATGACGAATGACAGCGGAGCTAGAGACCCCTCCACGGCCTGACCTCGGCTG	960
QY	2743	CAGTAAAGTTACGTGAGGCGCTGTCTCTCGGGGCTTGAAGTGGAGCCATCAGTTGCTT	2802
Db	961	CAGTAAAGTTACGTGAGGCGCTGTCTCTCGGGGCTTGAAGTGGAGCCATCAGTTGCTT	1020
QY	2803	TGCTGACCCCTCGGAGCAAGGCGCGCAAGGTGTGCTGAGACAGCTGTGCGGGGGGC	2862
Db	1021	TGCTGACCCCTCGGAGCAAGGCGCGCAAGGTGTGCTGAGACAGCTGTGCGGGGGGC	1080
QY	2863	CCCAAGCTGCGCGGGCTCCAGGCCCAACCAAGCTGTGCTGAGAGTCAAGGCCAATCTCC	2922
Db	1081	CCCAAGCTGCGCGGGCTCCAGGCCCAACCAAGCTGTGCTGAGAGTCAAGGCCAATCTCC	1140
QY	2923	CAGCACTGTGATCTGAGTTACGCGTTAAGAACTCTCTCTGTGTTTGAAGACAGTTC	2982
Db	1141	CAGCACTGTGATCTGAGTTACGCGTTAAGAACTCTCTCTGTGTTTGAAGACAGTTC	1200
QY	2983	GGGTTGTTCATTTCTGTACATTTATCTCCATTTTTTAAAAAAGTTTCTCTGACGGCCC	3042
Db	1201	GGGTTGTTCATTTCTGTACATTTATCTCCATTTTTTAAAAAAGTTTCTCTGACGGCCC	1259
QY	3043	CACGGCCCGAGCGCGGTGAGCGTGTGTCATGAGCCTGGGGGCCCGGGCTTCGCGTGC	3102
Db	1260	CACGGCCCGAGCGCGGTGAGCGTGTGTCATGAGCCTGGGGGCCCGGGCTTCGCGTGC	1319
QY	3103	GCTCTGCGCAGGTGTCTTGGGACCCCATCTCTGCTTCAATTTCAGTGCAGCTGTGA	3162
Db	1320	GCTCTGCGCAGGTGTCTTGGGACCCCATCTCTGCTTCAATTTCAGTGCAGCTGTGA	1379
QY	3163	CAGAAAGCACTCACCAATTAACCTTTCTGAAAGCAAAAAAAGAAAAA 3213	
Db	1380	CAGAAAGCACTCACCAATTAACCTTTCTGAAAGCAAAAAAAGAAAAA 1430	

```

RESULT 6
US-10-108-260A-1143/C
; Sequence 1143, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20040005560A1e1 full length cDNA

```



Db	122	GCTTACCATCCAGACAGACAGAGACCCAGAGCTCAAGTACAGGCGGAAACCATCTGTT	181
Qy	191	TAACTTGTTCACACAGCTTATGAAGTGTTAAGTGAACCCCAACACAGGCGCATCTATGA	250
Db	182	TAACTTGTTCACACAGGCTTATGAATGTCTGTAGTGAACCCCAACACAGGCGCATCTATGA	241
Qy	251	TATATATGGGAAGAGAGACTCGAATATGGAAGATGGAGAGTTGTGTGAAGAGAGAGAAC	310
Db	242	TATATATGGGAAGAGAGAGACTCGAATATGGAAGATGGAGAGTTGTGTGAAGAGAGAGAAC	301
Qy	311	CCCTGCTGAAATTCGAGAGAGTTTGAAGCGGCTGCAGAGAGAGAGAGAGAGAGATT	370
Db	302	CCCTGCTGAAATTCGAGAGAGTTTGAAGCGGCTGCAGAGAGAGAGAGAGAGAGATT	361
Qy	371	GCAGCAGCGAACCAATCCCAAGGGAAAGATCAAGCGTTGAGATGATGCAACCGACTTTT	430
Db	362	GCAGCAGCGAACCAATCCCAAGGGAAAGATCAAGCGTTGAGATGATGCAACCGACTTTT	421
Qy	431	TGATCGCTATGATGAGAGATATGAAAGTGTCTCCGACAGTACTTTCCGACAGATTGAAT	490
Db	422	TGATCGCTATGATGAGAGATATGAAAGTGTCTCCGACAGTACTTTCCGACAGATTGAAT	481
Qy	491	TAATATAATGACATATCCAGTCCATTGAGGCAACCTTGAAGGAGACAGACAGCAT	550
Db	482	TAATATAATGACATATCCAGTCCATTGAGGCAACCTTGAAGGAGACAGACAGCAT	541
Qy	551	CCCTCTGGAAGGCTCTCAACCCAGAGTGAATAAGAGAGGTTTCATTTAACTTTGCGCT	610
Db	542	CCCTCTGGAAGGCTCTCAACCCAGAGTGAATAAGAGAGGTTTCATTTAACTTTGCGCT	601
Qy	611	CAGACGAGTAACTTCGGCAAAAGGAGTGGGAGAGATTGGATTGGAGCTGGAGACTTACA	670
Db	602	CAGACGAGTAACTTCGGTAAAGGAGTGGGAGAGATTGGATTGGAGCTGGAGACTTACA	661
Qy	671	GGGGCCCTTGTGTGGGTCTCAAGCTGTTCGGTATCTCACCAAGATGCTTTGTGACAAC	730
Qy	662	GGGGCCCTTGTGTGGGTCTCAAGCTGTTCGGTATCTCACCAAGATGCTTTGTGACAAC	721
Db	731	AAACTGTGCTCTGAGATTTCATCCCGTGAATTCGACCCGCGCTGACACATGTCTTAGC	790
Qy	722	AAACTGTGCTCTGAGATTTCATCCCGTGAATTCGACCCGCGCTGACACATGTCTTAGC	781
Db	791	TCGGAGCTTGAACAGAGAACCGGTGGGCTTACCTGACGTGGCGATGGGGTATCTCAATCAGC	850
Db	782	TCGGAGCTTGAACAGAGAACCGGTGGGCTTACCTGACGTGGCGATGGGGTATCTCAATCAGC	841
Qy	851	CATGAACACTAGCATGTCTCCGAGACACTTAAACACAGCCACTTCATCTGTGGCCCTGACGCT	910
Db	842	CATGAACACTAGCATGTCTCCGAGACACTTAAACACAGCCACTTCATCTGTGGCCCTGACGCT	901
Qy	911	GGGAATCCCTCACTCTCTTTCGACTGATCAGCTATCAGCAAAATCCAAAGATGACGATCA	970
Db	902	GGGAATCCCTCACTCTCTTTCGACTGATCAGCTATCAGCAAAATCCAAAGATGACGATCA	961
Qy	971	GACTCGTGTGAAGAGATCCCTCAAGCAGGCTTCTTTGGAGCGGTGTGGAGT 1023	
Db	962	GACTCGTGTGAAGAGATCCCTCAAGCAGGCTTCTTTGGAGCGGTGTGGAGT 1014	
RESULT 8			
US-10-037-270-820			
; Sequence 820, Application US/10037270			
; Publication No. US20030104529A1			
; GENERAL INFORMATION:			
; APPLICANT: Tang, Y. Tom			
; APPLICANT: Liu, Chenghua			
; APPLICANT: Asundi, Vinod			
; APPLICANT: Zhang, Jie			
; APPLICANT: Ren, Feiyun			
; APPLICANT: Chen, Rui-hong			
; APPLICANT: Zhao, Qing A.			
; APPLICANT: Weinman, Tom			

Query Match	25.9%	Score 832.2	DB 15	Length 1767
Best Local Similarity	99.1%	Pred. No. 3.8e-251		
Matches 837	Conservative 0	Mismatches 8	Indels 0	Gaps 0
1	GAAGGTTTCGCAAGATGCGCGCTTTCGACCGAGAGAGCTGACAAATGAAATCTATT	60		
139	GAAGGTTTCGCAAGATGCGCGCTTTCGACCGAGAGAGCTGACAAATGAAATCTATT	198		
61	ACTGTTGCTGAAGTGTGGGAGGAGGCTTTCGAAAGCTGAAAGCTTACCGGA	120		
199	ACTGTTGCTGAAGTGTGGGAGGAGGCTTTCGAAAGCTGAAAGCTTACCGGA	258		
121	GGCTCTGATGCTTACCATTCAGACAGAGACCCAGAGCTCAAGTCACAGCGG	180		
259	GGCTCTGATGCTTACCATTCAGACAGAGACCCAGAGCTCAAGTCACAGCGG	318		
181	AAGCACTGTTTAACTTGTTCACAGGCTTATGAAAGTGTGACCCCAACAGGG	240		
319	AAGCACTGTTTAACTTGTTCACAGGCTTATGAAAGTGTGACCCCAACAGGG	378		
241	CCATCTATGATATATATGAGAAAGAGACTTGGAAATGGAAGATGGAGGTTTGGAAA	300		
379	CCATCTATGATATATATGAGAAAGAGACTTGGAAATGGAAGATGGAGGTTTGGAAA	438		
301	GGAGAGAAACCCCTGCTGAATTTGAGAGAGATTGACGGCTCAGAGAGAGAGAAAG	360		
439	GGAGAGAAACCCCTGCTGAATTTGAGAGAGATTGACGGCTCAGAGAGAGAGAAAG	498		
361	AGAGAGATTCAGACAGCGAAACAATCCCAAGGGAACGATCAGCGTTGAGTAGTGCA	420		
499	AGAGAGATTCAGACAGCGAAACAATCCCAAGGGAACGATCAGCGTTGAGTAGTGCA	558		
421	CCGACCTTTTGAATGCTATGATGAGAGATGTAAGATGTGTCCGGCAGTAGCTTTCCG	480		
559	CCGACCTTTTGAATGCTATGATGAGAGATGTAAGATGTGTCCGGCAGTAGCTTTCCG	618		
481	AGATTGAAATTAATAAATGACATATTCAGAGCTTGAAGGCAACCTTGAACAGCAG	540		
619	AGATTGAAATTAATAAATGACATATTCAGAGCTTGAAGGCAACCTTGAACAGCAG	678		
541	ACACAGCCATCTCTCTGGAAGCCTTCGAACCCAGAAATGAGAAATGAGAGGTTTCATTA	600		







OTHER INFORMATION: Clone S3 + 125 DNA Sequence  
NAME/KEY: misc.feature  
LOCATION: (1)\_(844)  
OTHER INFORMATION: n Can be any nucleotide  
US-09-927-738-16

Query Match 14.2%; Score 456.8; DB 9; Length 844;  
Best Local Similarity 96.9%; Pred. No. 1e-132;

Matches 498; Conservative 0; Mismatches 12; Indels 4; Gaps 3;

QY 1350 CAGGAATCTGTCCGAAAGTAATTGAGCAGAAAGTCCAGAAATGGCTCATCTGC 1409  
DB 1 CAGGAATCTGTCCGA-GATAATTGAGCAGAAAGTCCAAATGGGCTCATCTGC 59  
QY 1410 AATGCTGTGTCGGAAGTTGTCAATGACAGAGCAGAAAGACGAGAAAGTGAAGT 1469  
DB 60 AATGCTGTGTCGGAAGTTGTGTCAATGACAGAGCAGAAAGACGAGAAAGTGAAGT 119  
QY 1470 AATGAGTGAAGTGTGCTGCTGAGTGTGGAAGCAGTGAAGTCTCATCGAG 1529  
DB 120 AATGAGTGAAGTGTG-CCTGCAAGTGTGCTGAGTGAAGCAGTGAAGTCTCATCG--A 176  
QY 1530 GCTTCAAGGCTGGGCTGCTGCTTTTATGACCCGTGTGTGGGGAAGAAAGACCTG 1589  
DB 177 GCTTCAAGGCTGGGCTGCTGCTTTTATGACCCGTGTGTGGGGAAGAAAGACCTG 236  
QY 1590 AAAATGCTGTATGTTTCCGGGGGCTCTGATCAGGTGATGTGTGACAGTGAAGCC 1649  
DB 237 AAAATGCTGTATGTTTCCGGGGGCTCTGATCAGGTGATGTGTGACAGTGAAGCC 296  
QY 1650 CTCCGATACCAAGCAGTCCCAAGATCATACAGTGAATTAATGCCAAGACGAG 1709  
DB 297 CTCCGATACCAAGCAGTCCCAAGATCATACAGTGAATTAATGCCAAGACGAG 356  
QY 1710 ATTTTAAAGGCCGCAAAAATCTTTTCTGGAGTCTACAAATTTGAAATGAAAAA 1769  
DB 357 ATTTTAAAGGCCGCAAAAATCTTTTCTGGAGTCTACAAATTTGAAATGAAAAA 416  
QY 1770 CCCGACATCAGATGTTTATTTTATTTATTTATTTATGAAAGTGTGACATTTCAAT 1829  
DB 417 CCCGACATCAGATGTTTATTTTATTTATTTATTTATGAAAGTGTGACATTTCAAT 476  
QY 1830 TATGTGAAGGACATGACACACCCGACTTTG 1863  
DB 477 TATGTGAAGGACATGACACACCCGACTGTG 510

## RESULT 11

US-09-884-441-326/c  
Sequence 326, Application US/09884441  
Patent No. US20020119158A1  
GENERAL INFORMATION:  
APPLICANT: Algate, Paul A.  
APPLICANT: Carter, Darick  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
FILE REFERENCE: 210121.462C7  
CURRENT APPLICATION NUMBER: US/09/884,441  
NUMBER OF SEQ ID NOS: 489  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 326  
LENGTH: 455  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-884-441-326

Query Match 14.2%; Score 455; DB 9; Length 455;  
Best Local Similarity 100.0%; Pred. No. 2.5e-132;

Matches 455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1074 AGCGTTGAGTTCCACAGGGCGTTTCTCAAAAGTCAAGCTCAACAGGGCGATCAGACA 1133  
|||||

DB 455 AGCGTTGAGTTCCACAGGGCGTTTCTCAAAAGTCAAGCTCAACAGGGCGATCAGACA 396

QY 1134 TACTTCTTCCCTATTTCATTTGACGACCAAGCTTGTGCGCCAGGCCATGTTCTATGCCACC 1193

DB 395 TACTTCTTCCCTATTTCATTTGACGACCAAGCTTGTGCGCCAGGCCATGTTCTATGCCACC 336

QY 1194 GTGGGGCTCTAGGCTCTACTTTTCCATCACCCTGTGATCATCAAAACATACCTCAGG 1253

DB 335 GTGGGGCTCTAGGCTCTACTTTTCCATCACCCTGTGATCATCAAAACATACCTCAGG 276

QY 1254 GCTCAGAAAGAGAAATTTGAGAAAGCAGAGGAAAGCCGCCACCGATGTGCTGACAG 1313

DB 275 GCTCAGAAAGAGAAATTTGAGAAAGCAGAGGAAAGCCGCCACCGATGTGCTGACAG 216

QY 1314 AAGAGCAGAGGCGGAGTCCGCTGCTGCTGATGCAAGAAATCTGTCCAGAGATTAAT 1373

DB 215 AAGAGCAGAGGCGGAGTCCGCTGCTGCTGATGCAAGAAATCTGTCCAGAGATTAAT 156

QY 1374 GAGGCAAGAGTCCAGAAATGGGCTCATCATGTCAATGCCGTGACGGAGTTGTC 1433

DB 155 GAGGCAAGAGTCCAGAAATGGGCTCATCATGTGTCAATGCCGTGACGGAGTTGTC 96

QY 1434 AATGACAAAGCAGAGAAAGCAGAAAGTGAAGTGAATGACGTGACGTGCTGACAG 1493

DB 95 AATGACAAAGCAGAGAAAGCAGAAAGTGAAGTGAATGACGTGACGTGCTGACAG 36

QY 1494 TGCTGTGAAGACTTGAAAGTCTCATCTCAGCGA 1528

DB 35 TGCTGTGAAGACTTGAAAGTCTCATCTCAGCGA 1

## RESULT 12

US-09-907-969-326/c  
Sequence 326, Application US/09907969  
Publication No. US20030091580A1  
GENERAL INFORMATION:  
APPLICANT: Mitcham, Jennifer L.  
APPLICANT: King, Gordon E.  
APPLICANT: Algate, Paul A.  
APPLICANT: Fling, Steven P.  
APPLICANT: Retter, Marc W.  
APPLICANT: Fanger, Gary Richard  
APPLICANT: Reed, Steven G.  
APPLICANT: Vedvick, Thomas S.  
APPLICANT: Carter, Darick  
APPLICANT: Hall, Paul  
TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER  
FILE REFERENCE: 210121.462C8  
CURRENT APPLICATION NUMBER: US/09/907,969  
NUMBER OF SEQ ID NOS: 596  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 326  
LENGTH: 455  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-907-969-326

Query Match 14.2%; Score 455; DB 10; Length 455;  
Best Local Similarity 100.0%; Pred. No. 2.5e-132;

Matches 455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1074 AGCGTTGAGTTCCACAGGGCGTTTCTCAAAAGTCAAGCTCAACAGGGCGATCAGACA 1133

DB 455 AGCGTTGAGTTCCACAGGGCGTTTCTCAAAAGTCAAGCTCAACAGGGCGATCAGACA 396

QY 1134 TACTTCTTCCCTATTTCATTTGACGACCAAGCTTGTGCGCCAGGCCATGTTCTATGCCACC 1193

DB 395 TACTTCTTCCCTATTTCATTTGACGACCAAGCTTGTGCGCCAGGCCATGTTCTATGCCACC 336

QY 1194 GTGGGGCTCTAGGCTCTACTTTTCCATCACCCTGTGATCATCAAAACATACCTCAGG 1253

Db	335	GTGGGGCCCTTAGTGGTCTACTTTGGCATTGCACCGTGTGATTCATCAACCACTACTCAGG	276
Qy	1254	GCTCAAGAAAGGAAGGAATTGGAGAGCAGAGGGGAAAGCCGCCACCGATGTGCTGCAG	1313
Db	275	GCTCAGAAAGGAAGGAATTGGAGAGCAGAGGGGAAAGCCGCCACCGATGTGCTGCAG	216
Qy	1314	AAGAAGCAGAGGGCGGAGTCCGCTGTCCGGCTGATGCAAGAAATCGTCCGAGAGTAATT	1373
Db	215	AGAGAGCAGAGGGCGGAGTCCGCTGTCCGGCTGATGCAAGAAATCTGTCCGAGAGTAATT	156
Qy	1374	GAGCAGAGAAGTCCAGAAATGGGCTCATCTATCGTCATATGCCGTGTACGGAGAGTTGTCTC	1433
Db	155	GAGCAGAGAAGTCCAGAAATGGGCTCATCTATCGTCATATGCCGTGTACGGAGAGTTGTCTC	96
Qy	1434	AATGACAAAGCAGAGAGAGCCGAGAAGGTGAAGGTGAATTCGATCTGTGCTCCCTCTGCAG	1493
Db	95	AATGACAAAGCAGAGAGAGCCGAGAAGGTGAAGGTGAATTCGATCTGTGCTCCCTCTGCAG	36
Qy	1494	TGCCGTGGTGAAGGACTGCAAGCTCATCTCTCAGGGA	1528
Db	35	TGCCGTGGTGAAGGACTGCAAGCTCATCTCTCAGGGA	1

```

RESULT 13
US-09-827-271-326/c
; Sequence 326, Application US/09827271
; Publication No. US20030165504A1
; GENERAL INFORMATION:
; APPLICANT: Retter, Marc W.
; APPLICANT: Fanger, Gary R.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.462C6
; CURRENT APPLICATION NUMBER: US/09/827,271
; CURRENT FILING DATE: 2001-04-04
; NUMBER OF SEQ ID NOS: 461
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 326
; LENGTH: 455
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-827-271-326

```

Query Match	14.2%;	Score 455;	DB 10;	Length 455;
Best Local Similarity	100.0%;	Pred. No. 2.5e-132;		
Matches 455;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

Qy	1074	AGCGTTGGAGTTTCCACAGGAGCGTTTCTCTCAAAAGTCAAGTCTCAACAGGAGCCAGTCAACA	1133
Db	455	AGCGTTGGAGTTTCCACAGGAGCGTTTCTCTCAAAAGTCAAGTCTCAACAGGAGCCAGTCAACA	396
Qy	1134	TACTTCTTCCCTATTCACCTTACGAGCAACAGGTTCTGACCACAGGCAATGTTATGCAACC	1193
Db	395	TACTTCTTCCCTATTCACCTTACGAGCAACAGGTTCTGACCACAGGCAATGTTATGCAACC	336
Qy	1194	GTGGGGCCTCTAGTGTGTCTACTTTTGCCATGCAACCGTCTGATCATCAACCATACCTGAGG	1253
Db	335	GTGGGGCCTCTAGTGTGTCTACTTTTGCCATGCAACCGTCTGATCATCAACCATACCTGAGG	276
Qy	1254	GCTCAAGAAAGAAAGAAATTTGAGAGACGAGGAGAAAGCCGCCACCGATGTGCTCGAG	1313
Db	275	GCTCAAGAAAGAAAGAAATTTGAGAGACGAGGAGAAAGCCGCCACCGATGTGCTCGAG	216
Qy	1314	AAGAAGCAAGAGGGGGAGTCCGCTGTCGCTGATGACAGAAATCTGTCGGAAGATATT	1373
Db	215	AAGAAGCAAGAGGGGGAGTCCGCTGTCGCTGATGACAGAAATCTGTCGGAAGATATT	156
Qy	1374	GAGGCAAGAGTCCAGAAATGGGCTCATCTCATGCTCAATGCGTGTAAAGGAAATTTGTC	1433
Db	155	GAGGCAAGAGTCCAGAAATGGGCTCATCTCATGCTCAATGCGTGTAAAGGAAATTTGTC	96
Qy	1434	AATATCAAGACGACGAAGAGCGAGAAAGGTGAAGGTGATTCAGTGCATCTGTCGCCCTGACG	1493

Db 95 AATGACAAAGCGAGGAAGCCGAGGAGGATGATTGACGTGACTGTGCCCCCTGCAG 36

Qy 1494 TGCCTGATGAGGACTCGAAGCTCATCTTCAACGGA 1528

Dd 35 TGCCTGATGAGGACTCGAAGCTCATCTTCAACGGA 1

```

RESULT 14
US-10-198-053-326/c
: Sequence 326, Application US/10198053
: Publication No. US20030124140A1
: GENERAL INFORMATION:
: APPLICANT: Bangur, Chaitanya S.
: APPLICANT: Retter, Marc W.
: APPLICANT: Fanger, Gary R.
: APPLICANT: Hill, Paul
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
: TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
: FILE REFERENCE: 210121.462C9
: CURRENT APPLICATION NUMBER: US/10/198,053
: CURRENT FILING DATE: 2002-07-17
: NUMBER OF SEQ ID NOS: 624
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 326
: LENGTH: 455
: TYPE: DNA
: ORGANISM: Homo sapiens
: US-10-198-053-326

```

Query March 14.2% Score 455; DB 15; Length 455;  
Best Local Similarity 100.0%; Pred. No. 2.5e-12;  
Matches 455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1074	AGGCTTGGAACTTCCACAGGGGCGTTTCCTCAAAATCAAGCTCAACAGGGCCAGTCAGACA	1133
Db	455	AGCGTTGAGTTCACAGGGGCGTTTCCTCAAAATCAAGCTCAACAGGGCCAGTCAGACA	396
QY	1134	TACTTCTTCCCTATTCACCTTGACGGACCGCTTGCCACAGCGCATGTTCTATGCGCAC	1193
Db	395	TACTTCTTCCCTATTCACCTTGACGGACCGCTTGCCACAGCGCATGTTCTATGCGCAC	336
QY	1194	GTGGGGCTCTAGTGTGTCTACTTTGGCATGACCGGTCTGATCATCAACCTACTCTCAG	1253
Db	335	GTGGGGCTCTAGTGTGTCTACTTTGGCATGACCGGTCTGATCATCAACCTACTCTCAG	276
QY	1254	GCTCAGAAAAGAAAGAAATTGAGAAACAGAGGAAAAGCGCGCACCGATGTCTGCAG	1313
Db	275	GCTCAGAAAAGAAAGAAATTGAGAAACAGAGGAAAAGCGCGCACCGATGTCTGCAG	216
QY	1314	AAGAGCAAGAGGCGGAGTCCGCTGTCGGCTGATGCAAGATCTGTCCGAAGTAATT	1373
Db	215	AAGAGCAAGAGGCGGAGTCCGCTGTCGGCTGATGCAAGATCTGTCCGAAGTAATT	156
QY	1374	GAGGCACAAGAGTCCAAATAGGAGCTCATCATGTCATATGCTGATGCGGAGAGTTTGC	1433
Db	155	GAGGCACAAGAGTCCAAATAGGAGCTCATCATGTCATATGCTGATGCGGAGAGTTTGC	96
QY	1434	AATGACAAGACAGGAAGACGAGAGAGTGAAGTGAATTGACGTGACTGTGCCCTGCAG	1493
Db	95	AATGACAAGACAGGAAGAGGAGAGAGTGAAGTGAATTGACGTGACTGTGCCCTGCAG	36
QY	1494	TGCTGTGTAAGACTTGAAAGCTCATCTCTCACGGA	1528
Db	35	TGCTGTGTAAGACTTGAAAGCTCATCTCTCACGGA	1

RESULT 15  
US-09-764-868-167  
; Sequence 167, Application US/09764868  
; Patent No. US20020168712A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.

```
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT232
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 167
; LENGTH: 479
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (454)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (465)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-868-167
```

Query Match 14.1%; Score 452; DB 9; Length 479;

Best Local Similarity 98.1%; Pred. No. 2.3e-131;

Matches 471; Conservative 5; Mismatches 2; Indels 2; Gaps 2;

```
Qy 18 GCGACGGCTTGAGGAGAGAGAGCTGACATGAAAGACTATTACTGTTGCTGAAGCTG 77
Db 1 GCGACGGCTTGAGGAGAGAGAGCTGACATGAAAGACTATTACTGTTGCTGAAGCTG 60
Qy 78 GCGAGGAGGCTCTTCTGAAGAGCTGAAGCTGCTACCGGAGGCTCTGTATGCTTAC 137
Db 61 GCGAGGAGGCTCTTCTGAAGAGCTGAAGCTGCTACCGGAGGCTCTGTATGCTTAC 120
Qy 138 CATCCAGCAAGCAGACAGACCCAGAGCTCAAGTCAAGCGGAAAGAGCTGTTAACCTT 197
Db 121 CATCCAGCAAGCAGACAGACCCAGAGCTCAAGTCAAGCGGAAAGAGCTGTTAACCTT 180
Qy 198 GTTACCCAGGCTTATGAAGTGCTTAGTACCCCAACCAAGGGCCATGATATATAT 257
Db 181 GTTACCCAGGCTTATGAAGTGCTTAGTACCCCAACCAAGGGCCATGATATATATAT 240
Qy 258 GGGAGAGAGAGACTGGAATGGAAGATGGAAGGTTGTGAAGAGAGAAACCCCTGCT 317
Db 241 GGGAGAGAGAGACTGGAATGGAAGATGGAAGGTTGTGAAGAGAGAAACCCCTGCT 300
Qy 318 GAAATTGAGAGAGAGTTGAGCGGCTGCAAGAGAGAGAGAAAGAGAGATTGCAGCAG 377
Db 301 GAAATTGAGAGAGAGTTGAGCGGCTGCAAGAGAGAGAGAAAGAGAGATTGCAGCAG 360
Qy 378 CGAACCAATCCCAAGGGAAGATCAAGCTTGAAGTAGATGACACCGACCTTTTGATGCG 437
Db 361 CGAACCAATCCCAAGGGAAGATCAAGCTTGAAGTAGATGACACCGACCTTTTGATGCG 420
Qy 438 TATGATGAGAGATGAAAGATGTCCCGCAGTAGCTTT-CCGCAATTGAAATTAAATA 496
Db 421 TATGATGAGAGATGAAAGATGTCCCGCAGTAGCTTTCCCGCAATTGAAATTAAATA 479
```

Search completed: July 13, 2004, 18:19:52

Job time : 1509 secs

**This page Blank (uspio)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 13, 2004, 08:46:53 ; Search time 7858 Seconds  
(without alignments)  
12210.128 Million cell updates/sec

Title: US-09-787-678A-12

Perfect score: 3213

Sequence: 1 gaaagctgcgaagatgcgcg.....gaaagcagaaaaa 3213

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_estc1:\*  
9: gb\_estc1:\*  
10: gb\_estc2:\*  
11: gb\_estc3:\*  
12: gb\_estc4:\*  
13: gb\_estc5:\*  
14: gb\_estf:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gse\_hum:\*  
18: em\_gse\_inv:\*  
19: em\_gse\_pln:\*  
20: em\_gse\_vtc:\*  
21: em\_gse\_fun:\*  
22: em\_gse\_mam:\*  
23: em\_gse\_mus:\*  
24: em\_gse\_pro:\*  
25: em\_gse\_rtd:\*  
26: em\_gse\_phg:\*  
27: em\_gse\_vtl:\*  
28: gb\_gse1:\*  
29: gb\_gse2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1471.8	45.8	2926	11 AK053156	AK053156 Mus muscu
2	953.4	29.7	1033	12 BM909566	BM909566 AGENCOURT
3	927.4	28.9	1019	12 BM927425	BM927425 AGENCOURT
4	921	28.7	1005	9 AL552317	AL552317 AL552317

5	917	28.5	1027	13 BQ066706	BQ066706 AGENCOURT
6	914.8	28.5	972	9 AL525212	AL525212 AL525212
7	882.6	27.5	972	9 AL562459	AL562459 AL562459
8	876.4	27.3	1019	13 BQ061784	BQ061784 AGENCOURT
9	869	27.0	1025	12 BQ054445	BQ054445 AGENCOURT
10	863	26.9	1053	13 BQ063535	BQ063535 AGENCOURT
11	861.6	26.8	930	13 BQ058032	BQ058032 AGENCOURT
12	858	26.7	935	13 BQ050592	BQ050592 AGENCOURT
13	857.2	26.7	1017	13 BQ062383	BQ062383 AGENCOURT
14	856.6	26.7	1018	13 BQ063825	BQ063825 AGENCOURT
15	852.8	26.5	989	9 AL575646	AL575646 AL575646
16	852.2	26.5	1124	13 BQ061457	BQ061457 AGENCOURT
17	851.6	26.5	3844	11 AK089983	AK089983 Mus muscu
18	850.2	26.5	955	13 BQ0712401	BQ0712401 AGENCOURT
19	847	26.4	924	13 BQ030960	BQ030960 BQ030960
20	840.4	26.2	1201	9 AL582666	AL582666 AL582666
21	836.4	26.0	896	12 BQ065525	BQ065525 603064011
22	833.4	25.9	1012	13 BQ061825	BQ061825 AGENCOURT
23	833.2	25.9	869	12 BQ068897	BQ068897 AGENCOURT
24	827.8	25.8	997	13 BQ056258	BQ056258 AGENCOURT
25	825.8	25.7	1201	9 AL561468	AL561468 AL561468
26	822.4	25.6	990	12 BQ062756	BQ062756 AGENCOURT
27	821.6	25.6	962	13 BQ057649	BQ057649 AGENCOURT
28	820.4	25.5	987	13 BQ056732	BQ056732 AGENCOURT
29	816.6	25.4	823	12 BQ056757	BQ056757 602977620
30	816	25.4	891	12 BQ013882	BQ013882 603639484
31	815.4	25.4	969	13 BQ056772	BQ056772 AGENCOURT
32	812.4	25.3	825	9 AL040438	AL040438 DXF2p43C
33	810.6	25.2	969	13 BQ050638	BQ050638 BQ050638
34	810	25.2	887	13 BQ0187518	BQ0187518 AGENCOURT
35	807.4	25.1	904	13 BQ056088	BQ056088 AGENCOURT
36	806.6	25.1	934	13 BQ056939	BQ056939 AGENCOURT
37	803.8	25.0	949	13 BQ0711401	BQ0711401 AGENCOURT
38	800.8	24.9	920	13 BQ0646154	BQ0646154 AGENCOURT
39	800.6	24.9	1031	13 BQ062757	BQ062757 AGENCOURT
40	800	24.8	938	13 BQ0643192	BQ0643192 AGENCOURT
41	796.8	24.8	1201	13 BQ032157	BQ032157 BQ032157
42	796.4	24.8	964	13 BQ0642728	BQ0642728 AGENCOURT
43	792.8	24.7	1057	13 BQ063938	BQ063938 AGENCOURT
44	791.4	24.6	904	14 CA489230	CA489230 AGENCOURT
45	791.2	24.6	1001	13 BQ066712	BQ066712 AGENCOURT

## ALIGNMENTS

RESULT 1  
AK053156  
LOCUS  
DEFINITION Mus musculus 0 day neonate lung cDNA, RIKEN full-length enriched library, clone:R030019A03 product:hypothetical DnaJ N-terminal domain containing protein, full insert sequence.  
ACCESSION AK053156  
VERSION AK053156.1 GI:26343260  
KEYWORDS HTC; CAP trapper.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
REFERENCE  
1 Carninci, P. and Hayashizaki, Y.  
High-efficiency full-length cDNA cloning  
Meth. Enzymol. 303, 19-44 (1999)  
JOURNAL MEDLINE  
PUBMED 10349636  
AUTHORS  
2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes  
Genome Res. 10 (10), 1617-1630 (2000)  
JOURNAL MEDLINE  
PUBMED 11042159

REFERENCE  
AUTHORS  
3  
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,  
Komori, H., Akiyama, J., Nishii, K., Kikunishi, T., Tashiro, H., Itoh, M.,  
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,  
Yamamoto, R., Matsuno, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,  
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M.,  
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunaga, S., Kawai, J.,  
Ozaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system-384-format  
sequencing pipeline with 384 multicapillary sequencer  
Genome Res. 10 (11), 1757-1771 (2000)

JOURNAL  
MEDLINE  
20530913  
11076861

REFERENCE  
AUTHORS  
4  
The RIKEN Genome Exploration Research Group Phase II Team and the  
FANTOM Consortium  
Functional annotation of a full-length mouse cDNA collection  
Nature 409, 685-690 (2001)

JOURNAL  
AUTHORS  
5  
The FANTOM Consortium and the RIKEN Genome Exploration Research  
Group Phase I & II Team.  
Analysis of the mouse transcriptome based on functional annotation  
of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)

JOURNAL  
AUTHORS  
6  
(bases 1 to 2926)  
Fukuda, S., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,  
Akachi, J., Tanaka, K., Hanagaki, T., Hara, A., Hashizume, W.,  
Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T.,  
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, T., Kasukawa, T.,  
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Komori, H., Kouda, M.,  
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,  
Nakamura, M., Nishi, K., Nomura, K., Numata, R., Ohno, M., Ohsato, N.,  
Ozaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,  
Sano, H., Sasaki, D., Shibata, K., Shingawa, A., Shiraki, T.,  
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,  
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,  
Yamamatsu, M., and Hayashizaki, Y.  
Direct Submission  
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of  
Physical and Chemical Research (RIKEN), Laboratory for Genome  
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),  
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,  
Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp/  
URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,  
Fax: 81-45-503-9216)  
CDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in Riken.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.  
Please visit our web site for further details.  
URL: http://genome.gsc.riken.go.jp/  
URL: http://fantom.gsc.riken.go.jp/.  
Location/Qualifiers  
1. 2926  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="FANTOM,DB:E030019A03"  
/db\_xref="GI:2424521"  
/db\_xref="taxon:10090"  
/clone="E030019A03"  
/tissue\_type="lung"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="0 day neonate"  
16. 1695  
/note="unnamed protein product; hypothetical DnaU  
N-terminal domain containing protein (InterPro|IPR001623,  
evidence: InterPro)  
putative"  
/codon\_start=1  
/protein\_id="BAC35287.1"  
/db\_xref="GI:26343261"

FEATURES  
source

/translation="MATALSEBELNEDYSLNVRREASSELKAAYRLQMLYHPD  
KRPDELKQAKRFLVHQAYEVSDPQTRAIYIDYGRGEMEMEVKERTPAE  
IRSEFERLQREBERRIQORTPKPTISGVNATLPDYRDEVEDVSSGSPQIEIN  
KHHSIOSITAPITATDITALSSISTONONGGSYNFALRPTASKMELEIFGADL  
QSLPGLKLPRLNTRPCFVTLALQIPHSRLSRPGITVLANLDNRTGVIQMRGT  
QAMMTISIVRDKTRCHFYVALQIGIPHSRLSYHKEFDDQTRVKSLLKGFPTI  
VYGGAEKTSRSHSVGLAIVSISQPVGSIKVLNLSASQYFPPIHITDQLPSAVYA  
TQSPVAVYAVRLLIRPYLRQKEKLEKRENTASDIQKQEAFAVRLEQSEVR  
RIIEEESMGILIVNMYGKRVNKRREKRVKVIDVPLQCLVKSGLILTEASK  
ALPQEPDPCVCEKSLRVLYPQREVLYHWPVDBALRIRPQSRIDPDS"  
polyA\_signal  
2900..2905  
/note="putative"  
2926  
polyA\_site  
/note="putative"  
ORIGIN  
Query Match 45.8%; Score 1471.8; DB 11; Length 2926;  
Best Local Similarity 79.0%; Pred. No. 0;  
Matches 1917; Conservative 0; Mismatches 437; Indels 73; Gaps 11;  
QY 1 GAAAGGTGCGAAGATGCGAGCGGCTTGAGGAGGAGGAGGCTGACATTAAGACTATT 60  
DB 2 GAAAGGTGCGAAGATGCGAGCGGCTTGAGGAGGAGGAGGCTGACATTAAGACTATT 61  
QY 61 ACTGCTTGCGAAGCGCGCAGGAGGCTCTTCTTGAAAGCTGAAAGCTGCTACCGGA 120  
DB 62 ACTCACTGCTGAACGTCGCCAGGAGGCTCTTCTTGAAAGCTGAAAGCTGCTACCGGA 121  
QY 121 GGTCTGTATGCTTACATCCACAGACACAGACACCAAGCTCAAGTCAACGCGG 180  
DB 122 GGTGTGTATGCTTACATCCGCGACACAGACACCAAGCTCAAGTCAACGCGG 181  
QY 181 AACGATGTTTAACTGTTTCAACGAGCTTATGAAGTGTGTTGTCACCCCAACAGG 240  
DB 182 AGCGGTGTTTAACTGTTTCAACGAGCTTATGAAGTGTGTTGTCACCCCAACAGG 241  
QY 241 CCATCTATGATATATATGGAAGAGAGAGCTGAAATGGAAGAGAGTGTGGA 300  
DB 242 CTATCTATGATATATATGGAAGAGAGAGCTGGAAGATGGAAGATGGAAGTGTGGA 301  
QY 301 GGAAGGAACCCCTGCTGAAATTCAGAGAGAGTGTGAGCGCTGCAGAGAGAGAG 360  
DB 302 GAAAGGAACCCCGAGCGGATCAGAGAGAGTGTGAGACTCCAGAGAGCGGAG 361  
QY 361 AGAGAGATTTGAGAGCGGACCAATCCCAAGGAAAGATCAAGCTTGAAGATGCA 420  
DB 362 AGAGAGATTTGAGAGCGGACCAATCCCAAGGAAAGATCAAGCTTGAAGATGCA 421  
QY 421 CCGACCTTTTGTATCGCTATGATGAGAGATGAGATGTGTCGACAGTATGCTTCCG 480  
DB 422 CTGACCTTTTGTATCGCTATGATGAGAGATGAGATGTGTCGACAGTATGCTTCCG 481  
QY 481 AGATTGAATTAATAAATGCAATATCCAGTCCATTGAGGACCTTGAAGAGAG 540  
DB 482 AGATTGAATTAATAAATGCAATATCCAGTCCATTGAGGACCTTGAAGAGAG 541  
QY 541 AACAGGCAATCTCTCTGGAAGCTCTCAACCAAGATGGAATGGAAGAGGTTCCATTA 600  
DB 542 AACAGGCAATCTCTCTGGAAGCTCTCAACCAAGATGGAATGGAAGAGGTTCCATTA 601  
QY 601 ACTTTGCGCTCAGACGATTAATTCGCAAGAGAGTGGAGAGATTTGAGAGCTG 660  
DB 602 ACTTTGCACTCGAGAGTGTCACTCTGCAAGAGAGTGGAGAGATTTGAGAGCTG 661  
QY 661 GAGACCTCAAGGAGCTTTGTTGCTGAGCTGCAAGCTGTTCCGTATCTCACAACAAGTGT 720  
DB 662 GAGACCTCAAGGAGCTTTGTTGCTGAGCTGCAAGCTGTTCCGTATCTCACAACAAGTGT 721  
QY 721 TTGTGCAACAACGTCGTGAGGTTTCAATCCGCTGGAATTCGACCCGCGCTGAGCA 780  
DB 722 TTGTGCAACAACGTCGTGAGGTTTCAATCCGCTGGAATTCGACCCGCGCTGAGCA 781  
QY 781 CTGTCTAGCTGGAACCTAGACAAGAACACGCTGGCTAAGCTGAGTGGGCTA 840

```
Db      782 CTGTGCTAGCTCGGAACTTGGACAAAGAACTGTGGGCTACTGTGAGTGGGGGCTA 841
Qy      841 TCCAGTCAGCCATGAACTAGATCGTCCGAGACATTAACAGCCACTTCACTGTGG 900
Db      842 TCCAGTCAGCCATGAACTAGATCGTCCGAGACATTAAGACCTGCACTTCACTGTGG 901
Qy      901 CCTGAGCTGGGAACTCCCTCACTCTTGGACATGATCAGTATCAGCAAAATTTCCAG 960
Db      902 CCTGAGCTGGGATCCCTCACTCTTGGACATGATCAGTATCAGCAAAATTTCCAG 961
Qy      961 ATGACATCAGACTCGTGTGAAGATCCCTCAAGAGGCTTCTTGGAGCGTGTGG 1020
Db      962 ACAGACACAGACTCGAGTGAAGGTCCTCAAGGAGGCTTCTTGGAGCACTATGG 1021
Qy      1021 AGTACGAGCTGAGAGAAATCTCCAGGACACGCTTTTGGTGTGAGCTGTGAGCTGG 1080
Db      1022 AGTATGAGCTGAGAGAAATTTCCGACACAGCGCTTGGGTGGGCTGTGAGCAATTG 1081
Qy      1081 GAGTTCCACAGGGGCTTCTCTCAAGTCAAGTCAAGGAGCCAGTCAATCTTCT 1140
Db      1082 GCGTCCCTCAGGGGTGTCTCTCAAGTCAAGTCAAGGAGCCAGTCAAGCTTCTTCT 1141
Qy      1141 TCCCTATTCACTTGAACGAGACGCTTGGCCAGGCGCATGTCTATGACACGCTGGAGC 1200
Db      1142 TCCCATTCACACTGACGAGTACGCTGCTCCGAGGCGTGTCTTACGACCGTGGGGC 1201
Qy      1201 CTCTAGTGTCTACTTTTGCATGACGCTGTGATCATCAACATACCTCAGGCTCAGA 1260
Db      1202 CGCTGTGTCTACTTTTGCATGACGCTGTGATCATCAAGCTTCTCAGGCTCAGA 1261
Qy      1261 AAGAGAGAAATTTGAGAAAGAGAGGAAAGCGGCGCACCGAGTGTCTGAGAAAGC 1320
Db      1262 AAGAGAGAAATTTGAGAAAGAGAGAAAGAGAGAAAGAGAGAGATCTGAGAAAGC 1321
Qy      1321 AAGAGAGGAGTCCGCTGTCCGAGCTGATGACAGATCTGTCCAGAGATTAATGAGCAG 1380
Db      1322 AAGAGAGAGAGCGCGCTGTCCGAGTATGACAGATCTGTCCAGAGATTAATGAGCAG 1381
Qy      1381 AAGAGTCCAGANTGGGCTCTCATCTCATGCTCAATGCTGTGATGAGGAGTTTGTCAATGACA 1440
Db      1382 AAGAGTCCAGANTGGGCTCTCATCTCATGCTCAATGCTGTGATGAGGAGTTTGTCAATGACA 1441
Qy      1441 AGAGCAGAGAGAGAGAGAGAGTGAAGTGAATGACGTGTGCTCCGAGTGTGCTGG 1500
Db      1442 AGAGCAGAGAGAGAGAGAGAGTGAAGTGAATGACGTGTGCTCCGAGTGTGCTGG 1501
Qy      1501 TGAAGACTCGAAGCTCATCTCTCAGGAGGCTCCAAAGGCTGGCTGGCTGCTTTTATG 1560
Db      1502 TGAAGACTCGAAGCTCATCTCTCAGAGGCTCCAAAGGCTGGCTGGCTGGCTTTATG 1561
Qy      1561 ACCGTGTGTGGGAGAGAGAGAACTGAAGTGTCTATCAGTTCCGGGGCTGCTGC 1620
Db      1562 ATCCATGTGTGGGAGAGAGAGAGCTGCGAGTGTGATCAAGTTCCGAGTGTGCTGC 1621
Qy      1621 ATCAGGTGATGTGCTGAGAGTGAAGGCTCCCGGATTAACAAAGCAATGCCACAGATGG 1680
Db      1622 ACCAGGTATGTGCTGAGAGTGAAGGCTCCGAGATTAACAAAGCAATGCCACAGATGG 1681
Qy      1681 ATACAGATGATTAATGCTG-C-CAAGAACAGATTTTAAAGGCGCAAAATTTTCC 1739
Db      1682 ACACAGACGGGTAAACCGCTTGGGAACTCGAGAGAAAGAGGCTAGAGAAATCTTTCC 1741
Qy      1740 TGGAGTCTACAAATTTGGAA-ATGAAAAAACCCAGACATCAGATGTTTTTATTTAT 1798
Db      1742 TAAAGAGCTCAAAATTTGGAAGCAGGAAAAACCCAGACATGAGATGTTTTTATTTAT 1801
Qy      1799 TATATATATGAAGAGTGTACATTAATTAATGTAAGGAGATGACACCCAGC 1858
Db      1802 CTCA-----GAAAGGGGGCAGCTGTATCAATTAATTAAGGAGATGACACCCAGC 1855
Qy      1859 TTTTGAAGGTCTGTGGGGTGAAGCTGAGGAGCCCACTGGGAAACAGACTGACCTGG 1918
```

```
Db      1856 TCCATGAGTGTGT-GGGGTGAAGTGAAGAGCCCTGTGGCCAGACCCAGACGCGCATG 1914
Qy      1919 CCATGAGCTTTTTCACAGATCATGTTCTGTGAAGGAGGCTGTGGCCCTGATCCGC 1978
Db      1915 TC-----TTCGCCAAGATCATGTCTGTGAAGGTCAGGCCCTGTGTCTT----- 1959
Qy      1979 TGTGTCCGAGCAACAGTGTGACCGAGCCCGCCCTGTAGTTCTTGGCTGGGTCTG 2038
Db      1960 -----CTCATTAAGTGTGTGTGACAGACGCTTCTC 1988
Qy      2039 GAGGTGTGTGTGAGACACCTGTGCTTCAACAGAGAGCGTGAACATTTGTGAGTCCAC 2098
Db      1989 TGAAGGATTTGGAATATGCTGTGCA---CAGATGGAGCTGCTTCTGTAAGCCCAT 2045
Qy      2099 GCTGAACATGGAACAACTGAAAAAGCAGGCGCTCCCGGTCAAGGAGCTGTGCTG 2158
Db      2046 GCTGTGTGTGGGAAACATGGAAGAAAGAGAGAGTGTGTG-----TGTGCTC 2095
Qy      2159 TGTGTGCTTCCATGA-CCACCTCTCTGTGTAATATTAATCTGTAATCTGAGCAG 2217
Db      2096 AGTGAATCTTCCACAGCCACCTCTCTTCTTAAGACGTACATGAATGA-CTGAGAGG 2154
Qy      2218 ATTGCGGTTTATTAACATGCTTTTATCTGAGAAACAGGTTTGAATTAATGCTGTC 2277
Db      2155 ACTGCGGTTTATTAACATGCTTTTATCTGAGAAACATGTTGAAATTAATGCTGTC 2214
Qy      2278 TTTTTCCTCCATCTCCAGAGCTGCTCAATGATTCACCGGCCCCCTGAGCTGGAGAG 2337
Db      2215 TTTTTCCTCCATCTTATTAACAGAGCTTCTCAAAATCTTCTCCAGGCTTAACTGACAGTA 2274
Qy      2338 GTATGATTAATCTCCAGTCCAGAGGCTAGCCCTGACAGAGTGGCTTCCGTATCCCGG 2397
Db      2275 GGGGGGACAGCCTGTGCTCCAGG---TGCACTGTCCAGGACACATGCCCAATGCTTA 2331
Qy      2398 TGGAAAAACGCTCTGACACAGCGGCG 2424
Db      2332 AGAAGAGAGCCCTGTGCTGACAGTGAAC 2358
```

```
RESULT 2
BM909566 1033 bp mRNA linear EST 12-MAR-2002
LOCUS BM909566
DEFINITION AGENCOURT_6640947 NIH_MGC_99 Homo sapiens cDNA clone IMAGE:5434173
5', mRNA sequence.
ACCESSION BM909566
VERSION BM909566.1 GI:19359945
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1033)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@db-remail.nih.gov
Tissue Procurement: Lou Staudt
CDNA Library Preparation: Rubin Laboratory
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: L1CM1910 row: m column: 22
High quality sequence stop: 733.
Location/Qualifiers
1..1033
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5434173"
/tissue_type="lymphoma, cell line"
```





Db	121	GTATCTCTACCATCCAGCAAGACACAGACCCAGAGCTCAAGTCCACAGGGGAAACAC	180
Qy	187	TGTTTAACCTTGTTCACCGAGGCTTATGAAGTCTTAGTACCCCAACCAAGGGGCATCT	246
Db	181	TGTTTAACCTTGTTCACCGAGGCTTATGAAGTCTTAGTACCCCAACCAAGGGGCATCT	240
Qy	247	ATGATATATATGGGAAGAGAGGACTGGAAAATGAAAGATGGAGGTTGTGGAAAGAGGA	306
Db	241	ATGATATATATGGGAAGAGAGGACTGGAAAATGAAAGATGGAGGTTGTGGAAAGAGGA	300
Qy	307	GAAACCCCTGTGAAATTTGAGAGGACTTGAACCGCTGCAAGAGAGAGAAAGAGGA	366
Db	301	GAAACCCCTGTGAAATTTGAGAGGAGTTTGACCGCTGCAAGAGAGAGAAAGAGGA	360
Qy	367	GATTGCAGACGGAACCAATCCCAAGGGAAACATACAGCGTTGAGTATGTCACCGACC	426
Db	361	GATTGCAGACGGAACCAATCCCAAGGGAAACATACAGCGTTGAGTATGTCACCGACC	420
Qy	427	TTTTTGATGCTATGATGAGAGATGAAGAATGTGTCCGGCAGTAGCTTTCGACGATTG	486
Db	421	TTTTTGATGCTATGATGAGAGATGAAGAATGTGTCCGGCAGTAGCTTTCGACGATTG	480
Qy	487	AAATTAATTAATTCGACATATCCCAAGTCCATTGAGGACCCCTTGAACGGACAGACAG	546
Db	481	AAATTAATTAATTCGACATATCCCAAGTCCATTGAGGACCCCTTGAACGGACAGACAG	540
Qy	547	CCATCTCTCTGGAAGACCTCTCAACCCAGAAATGGAATGAGAGAGGTTCCATTAACTTGG	606
Db	541	CCATCTCTCTGGAAGACCTCTCAACCCAGAAATGGAATGAGAGAGGTTCCATTAACTTGG	600
Qy	607	CGCTCAGACGAGTAACTTCGGCAAAAGGATGGGGAAGTGGAAATTTGAGAGCTGAGACC	666
Db	601	CGCTCAGACGAGTAACTTCGGCAAAAGGATGGGGAAGTGGAAATTTGAGAGCTGAGACC	660
Qy	667	TACAGGGGCTTTTGTTCGGTCTCAAGCTGTTCCGTAAATCTCACACCAAGATGCTTTGGA	726
Db	661	TACAGGGGCTTTTGTTCGGTCTCAAGCTGTTCCGTAAATCTCACACCAAGATGCTTTGGA	720
Qy	727	CAACAAACGTGCTCTGACAGTTTTCATCCCGTGAATCGAACCGGGCCGACACTGACC	786
Db	721	CAACAAACGTGCTCTGACAGTTTTCATCCCGTGAATCGAACCGGGCCGACACTGACC	780
Qy	787	TAGCTCGGAACCTTAGACAAGAACACCGTGGGCTACCTGAGT-GGCGATGGGGTATCCAG	845
Db	781	TAGCTCGGAACCTTAGACAAGAACACCGTGGGCTACCTGAGTGGGGTATCCAG	840
Qy	846	TCAGCGATGAAACATAGCATGCTCGGAGACACTAAACCAAGCACTTCACTGTGGCCCTG	905
Db	841	TCAGCGATGAAACATAGCATGCTCGGAGACACTAAACCAAGCACTTCACTGTGGCCCTG	900
Qy	906	CAGCTGGGAATCCCTCACTCCCTT-GCATGATTCAGCTATCAGC-ACAAATTTCCAAAGATG	963
Db	901	CAGCTGGGAATCCCTCACTCCCTTGGACGAGTACGCTATACGACAAATTTCCAAAGATG	960
Qy	964	ACGA-----TCAGACTCGTGTGAAGAGATCCCTCAAGACAGGCTTCTTGTGGACGCTG	1016
Db	961	ACGAATCCGGAATTCGTGTGGGAAGGATCCCTCAAGACAGGCTTCTTGTGTGGGG	1019
RESULT 4			
AL552317			
LOCUS	AL552317	1005 bp	mRNA linear EST 31-MAY-2003
DEFINITION	AL552317 Homo sapiens PLACENTA COR 25-NORMALIZED Homo sapiens CDNA		
ACCESSION	clone CS0D106YF21 5-PRIME, mRNA sequence.		
VERSION	AL552317		
KEYWORDS	AL552317.2 GI:31274132		
SOURCE	EST.		
ORGANISM	Homo sapiens (human)		
	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
REFERENCE	1 (baaes 1 to 1005)		

AUTHORS	Li, W.B., Gruber, C., Jeessee, J. and Polayes, D.				
TITLE	Full-length cDNA libraries and normalization				
JOURNAL	Unpublished (2001)				
COMMENT	On Feb 15, 2001 this sequence version replaced gi:12891103.				
	Contact: Genoscope				
	Genoscope - Centre National de Sequencage				
	Bp 191 9106 EVRY cedex - France				
	Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr				
	Library was constructed by Life Technologies, a division of				
	Invitrogen. This sequence belongs to sequence cluster 271.r For				
	more information about this cluster, see				
	http://www.genoscope.cns.fr/				
	cgi-bin/cluster.cgi?seq=CS0D1069CH11Qp1&cluster=271.r. Contact :				
	Feng liang Email : fliang@liletech.com URL :				
	http://fulllength.invitrogen.com/Invitrogen Corporation 1600				
	Faraday Avenue Genoscope sequence ID : CS0D1069CH11Qp1.				
FEATURES	Location/Qualifiers				
source	1. 1005				
	/organism="Homo sapiens"				
	/mol_type="mRNA"				
	/db_xref="taxon:9606"				
	/clone="CS0D1069Yp21"				
	/issue_type="PLACENTRA COT 25-NORMALIZED"				
	/clone_id="Homo sapiens PLACENTRA COT 25-NORMALIZED"				
	/note="1st strand cDNA was primed with a NotI-oligo (dT)				
	primer. Five prime end enriched, double-strand cDNA was				
	digested with Not I and cloned into the Not I and EcoR V				
	sites of the pCMVSPORT 6 vector. Library was normalized."				
ORIGIN					
Query Match	28.7%	Score 921	DB 9	Length 1005	
Best Local Similarity	99.5%	Pred. No. 1.2e-228			
Matches 932	Conservative 2	Mismatches 2	Indels 1	Gaps 1	
QY	19	CGACGGCCTTGAAGGAGAGAGCTGACATATGAAGACTATTACTGCTGCAAGCGC	78		
DB	68	CGACGGCCTTGAAGGAGAGAGCTGACATATGAAGACTATTACTGCTGCAAGCGC	127		
QY	79	CGACGGAGCGCTCTTGTGAAGGCTGAAAGTGCGCTACCGAGGCTGTATGCTTACC	138		
DB	128	CGACGGAGCGCTCTTGTGAAGGCTGAAAGTGCGCTACCGAGGCTGTATGCTTACC	187		
QY	139	ATCCAGACAAACAGACAGAGACCAGAGCTCAAGTCAACAGCGCGAAAGACTGTTAACTTG	198		
DB	188	ATCCAGACAAACAGACAGAGACCAGAGCTCAAGTCAACAGCGCGAAAGACTGTTAACTTG	247		
QY	199	TTCACCAAGGCTTATGAAGTGTCTTGTGTGACCCCAACCAAGGCGCATCTATATATATG	258		
DB	248	TTCACCAAGGCTTATGAAGTGTCTTGTGTGACCCCAACCAAGGCGCATCTATATATATG	307		
QY	259	GGAAGAGAGAGCTGGAATAATGGAAGATGGAAGATGTTGGAAGAAAGAGAGAAACCCCTGCTG	318		
DB	308	GGAAGAGAGAGCTGGAATAATGGAAGATGGAAGATGTTGGAAGAAAGAGAGAAACCCCTGCTG	367		
QY	319	AAATTGAGAGAGATTTGAGCGCGCTGACAGAGAGAGAGAAAGAGAGAGATTTGACGAGC	378		
DB	368	AAATTGAGAGAGATTTGAGCGCGCTGACAGAGAGAGAGAAAGAGAGATTTGACGAGC	427		
QY	379	GAACCAATCCCAAGAGAAAGCATGACGCTTGAGATGATGCCACCGACTTTTGTATGCT	438		
DB	428	GAACCAATCCCAAGAGAAAGCATGACGCTTGAGATGATGCCACCGACTTTTGTATGCT	487		
QY	439	ATGATGAGAGATTAAGAAATGATGTCGCGAGTACTCTTCCGCAATTGAATTAATATAA	498		
DB	488	ATGATGAGAGATTAAGAAATGATGTCGCGAGTACTCTTCCGCAATTGAATTAATATAA	547		
QY	499	TGCAATATCCCAATGATGAGGACACCTTGAAGAGAGACAGACAGCCATCCTCTCTG	558		
DB	548	TGCAATATCCCAATGATGAGGACACCTTGAAGAGAGACAGACAGCCATCCTCTCTG	607		
QY	559	GAAGCTTCAACCCAGATGAAATGAGAGAGGTTCCATTACTTTGCGCTCAGACGAG	618		
DB	608	GAAGCTTCAACCCAGATGAAATGAGAGAGGTTCCATTACTTTGCGCTCAGACGAG	667		

Qy	619	TAACTTCGGCAAAAGGAGATGGGGAGAGTTGGAAATTTGGAGCTGGAGAACTTACAGGGGGCTT	678
Db	668	TAACTTCGGCAAAAGGAGATGGGGAGAGTTGGAAATTTGGAGCTGGAGAACTTACAGGGGGCTT	727
Qy	679	TGTTTGGGTCTCAAGCTGTGTTCCGTAACTTCACACCAAGATGCTTTGTGAACAACTGTG	738
Db	728	TGTTTGGGTCTCAAGCTGTGTTCCGTAACTTCACACCAAGATGCTTTGTGAACAACTGTG	787
Qy	739	CTGTGCAGTTTTCATCCCGTGGAAATCCGACCCGGGCTGACCACTGTCTTAGTGGGAAC	798
Db	788	CTGTGCAGTTTTCATCCCGTGGAAATCCGACCCGGGCTGACCACTGTCTTAGTGGGAAC	847
Qy	799	TAGACAAAGACACCGTGGCTTACCTGTCAAGTGGCGATGGGGTATTCAGTCAAGCATGAACA	858
Db	848	TAGACAAAGACACCGTGGCTTACCTGTCAAGTGGCGATGGGGTATTCAGTCAAGCATGAACA	907
Qy	859	CTAGCATGTGCCAGACACTAAAACACACCACTTCACTGTGGCCCTGAGACT-GGGAATC	917
Db	908	CTAGCATGTGCCAGACACTAAAACACACCACTTCACTGTGGGCCCTGAGACTGGGGAAATC	967
Qy	918	CCTCACTCTTTTGCACTGATCAGCTATTCAGCACAAT	954
Db	968	CCTCACTCTTTTGCACTGATTAAGCTATTCAGCACAAT	1004

RESULT 5	LOCUS	DEFINITION	AGENCOURT	1027 bp	mRNA	linear	EST 02-APR-2002
B0066706	B0066706	5', mRNA sequence.	68689906	NIH_MGC_99	Homo sapiens	CDNA clone	IMAGE:5931375
ACCESSION	B0066706						
VERSION	B0066706.1			GI:19895752			

SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens

REFERENCE 1 (bases 1 to 1027)  
AUTHORS NIH-MGC <http://mgc.nhl.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.

Tissue Procurement: Lou Staudt  
 cDNA Library Preparation: Rubin Laboratory  
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.lnl.gov>  
 Plate: LNCM212 row: j column: 16  
 High quality sequence stop: 676.

## FEATURES

### Source

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5931375"
/tissue_type="lymphoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_99"
/name="Organ: lymph; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCGAG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
library."

```

Query Match	28.5%;	Score 917;	DB 13;	Length 1027;
Best Local Similarity	96.1%;	Pred. No. 1.4e-227;		
Matches 983; Conservative	0;	Mismatches 35;	Indels 5;	Gaps 4;

OY	1	TTGGGAGATGGCGACCGGCTTTGAGCGACGAGGAGCTGGACAATGAAAGACTATTAACGCT	66
OY	1	TTGGGAGATGGCGACCGGCTTTGAGCGAGGAGGAGCTGGACAATGAAAGACTATTAACGCT	60
OY	67	TGCTGAACGTGCGCAGGAGGAGCGCTCTTCTGTAGAAGCTGAAAGCTGCTTACCGAGGCTCT	126
Db	61	TGCTGAACGTGCGCAGGAGGAGGCGCTCTTCTGTAGAAGCTGAAAGCTGCTTACCGAGGCTCT	120
OY	127	GTAATGCTCTACCATCCAGACAAAGACACAGAACCCAGAGCTTCAATCAACGCGGAAAGAC	186
Db	121	GTAATGCTCTACCATCCAGACAAAGACACAGAACCCAGAGCTTCAATCAACGCGGAAAGAC	180
OY	187	TGTTTAAACCTTGTTCACACAGGCTTATAGAGTGCTTAATGACCCCCAAACAGGCGCATCT	246
Db	181	TGTTTAAACCTTGTTCACACAGGCTTATAGAGTGCTTAATGACCCCCAAACAGGCGCATCT	240
OY	247	ATGATATATATGGAGAAGAGAGACTGGAAATGAGAAGATGGAGAGCTTGTGAAANAGAGGA	306
Db	241	ATGATATATATATGGAGAAGAGAGACTGGAAATGAGAAGATGGAGAGCTTGTGAAANAGAGGA	300
OY	307	GAACCCCTGCTGAAATTCAGAGAGAGATTGTGAGCGGCTGCAGAGAGAGAGAGAAAGAGA	366
Db	301	GAACCCCTGCTGAAATTCAGAGAGAGATTGTGAGCGGCTGCAGAGAGAGAGAGAAAGAGA	360
OY	367	GATTGCAGCAGCAGAACCAATCCCAAGGGAACGATCAGCGTTGGAGTAGATGCCACGAGCC	426
Db	361	GATTGCAGCAGCAGAACCAATCCCAAGGGAACGATCAGCGTTGGAGTAGATGCCACGAGCC	420
OY	427	TTTTTGAATCGTATGATGAGAGAGTATGAATGTGTCCGCGAGTAGTCTTTCCGAGATTG	486
Db	421	TTTTTGAATCGTATGATGAGAGAGTATGAATGTGTCCGCGAGTAGTCTTTCCGAGATTG	480
OY	487	AAATTAATTAATATGCATATCCCAATGCCAGTCCATATTGAGGACCCCTTACAGCGCAGACACAG	546
Db	481	AAATTAATTAATATGCATATCCCAATGCCAGTCCATATTGAGGACCCCTTACAGCGCAGACACAG	540
OY	547	CCATCTCTCTGGAAGCGCTCTCAACCCAGATGGAATGAGAGGTTCCATTAACTTTG	606
Db	541	CCATCTCTCTGGAAGCGCTCTCAACCCAGATGGAATGAGAGGTTCCATTAACTTTG	600
OY	607	CGCTCAGACGAGTAACTTTCCGCAAAAGGATGGGAGAGTTGGAATTTGGAGCTGGAGACC	666
Db	601	CGCTCAGACGAGTAACTTTCCGCAAAAGGATGGGAGAGTTGGAATTTGGAGCTGGAGACC	660
OY	667	TACAGGGGCGCTTGTTCGGTCTCAAGCGTGTTCGGTAATCTCACACCAAGATGCTTTGTGA	726
Db	661	TACAGGGGCGCTTGTTCGGTCTCAAGCGTGTTCGGTAATCTCACACCAAGATGCTTTGTGA	720
OY	727	CAACAAACTGTGCTCTGCAGTTTTCATTCGGTGGAAATCCGACCCGGCTGACACCTGTCC	786
Db	721	CAACAAACTGTGCTCTGCAGTTTTCATTCGGTGGAAATCCGACCCGGCTGACACCTGTCC	780
OY	787	TAGCTCGGAACCTTGAACAAGAACACCGTGGGCTTCTGCAGT--GGCGATGGGGTATATCA	844
Db	781	TAGCTCGGAACCTTGAACAAGAACACCGTGGGCTTCTGCAGTGGCGAATGGGGTATATCA	840
OY	845	GTCAGCCATG--AACACTGACATGCTGCCGAGCACT--AAAAACAGCCACTTCACTGTGGCC	902
Db	841	GTCAGCCATGAAACATGACATGCTGCCGAGCACTTAAAAACAGCCCACTTAACTGTGGCC	900
OY	903	CTGACGCTGGGAATCCCTCACTCC--TTTGAACGTATAGCTATACAGCAAAATTCAGAGA	961
Db	901	CTGACGCTGGGAATCCCTCACTCTTTTTCGCCGATAGCTATACAGCAAAATTCAGAGA	960
OY	962	TGACGATCAGACTGTGTGAAAGAGATCCCTCAAGCAGGCTTCTTTGGGACGGGTGTGGA	1021
Db	961	TGACGATCAGACTGTGTGAAAGAGATCCCTCAAGCAGGCTTCTTTGGGACGGGGGT	1020
OY	1022	GTA	1024

Db 1021 GGA 1023

RESULT 6  
AL525212 972 bp mRNA linear EST 22-MAY-2003  
LOCUS AL525212 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens  
DEFINITION cDNA clone CS0DC005YN12 5'-PRIME, mRNA sequence.

ACCESSION AL525212  
VERSION AL525212.2 GI:31043467  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished (2001)  
COMMENT On Feb 13, 2001 this sequence version replaced gi:12788705.  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 271.r For  
more information about this cluster, see  
http://www.genoscope.cns.fr/  
cgi-bin/cluster.cgi?seq=CS0DC005D606P1&cluster=271.r. Contact :  
Feng Liang Email : fliang@lifetech.com URL :  
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
Faraday Avenue Genoscope sequence ID : CS0DC005D606P1.  
Location/Qualifiers

FEATURES  
source 1..972  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0DC005YN12"  
/issue\_type="NEUROBLASTOMA COT 25-NORMALIZED"  
/clone\_1ib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"  
/note="1st strand cDNA was primed with a NotI-oligo (dt)  
primer. Five prime end enriched, double-strand cDNA was  
digested with Not I and cloned into the Not I and EcoR V  
sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match 28.5%; Score 914.8; DB 9; Length 972;  
Best Local Similarity 99.8%; Pred. No. 5e-227;  
Matches 916; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 11 GAAGATGGCGAGCGCTTTGAGCGAGAGAGCTGACATGAGACTTACTGCTGT 70  
55 GGATATGGCGAGCGCTTTGAGCGAGAGAGCTGACATGAGACTTACTGCTGT 114  
71 GAAGTGGCGAGCGCTTTGAGAGAGCTGAAAGCTCTTACCGGAGGCTCTGTAT 130  
115 GAAAGTGGCGAGGAGGCTCTTTGAGAGAGCTGAAAGCTCTTACCGGAGGCTCTGTAT 174  
131 GCTTACCATCAGACAGACAGAGAGCCAGAGCTCAAGTCAAGCGAGCGAGAGCTGT 190  
175 GCTTACCATCAGACAGACAGAGAGCCAGAGCTCAAGTCAAGCGAGCGAGAGCTGT 234  
191 TAACTTGTTCACAGGCTTATGAGAGTCTTAGTGAACCCCAACCAAGGCGCATCTATGA 250  
235 TAACTTGTTCACAGGCTTATGAGAGTCTTAGTGAACCCCAACCAAGGCGCATCTATGA 294  
251 TATATATGGAGAGAGAGAGCTGAAATGGAAGATGGAGGTTGTGAAAAGAGAGAAAC 310  
295 TATATATGGAGAGAGAGAGCTGAAATGGAAGATGGAGGTTGTGAAAAGAGAGAAAC 354  
311 CCCTGTGAAATTTGAGAGAGAGTGTGAGCGGCTGACAGAGAGAGAGAGAGAGATT 370  
|||||

Db 355 CCCTGTGAAATTTGAGAGAGAGTGTGAGCGGCTGACAGAGAGAGAGAGAGATT 414  
371 GCAGCAGAGAAACCAATCCCAAGGAGAGATCAGGCTTGGAGTAGATGCCACCGACTTTT 430  
415 GCAGCAGAGAAACCAATCCCAAGGAGAGATCAGGCTTGGAGTAGATGCCACCGACTTTT 474  
431 TGATCGCTATGATGAGAGAGATGAGATGATGATGATGATGATGATGATGATGATGAT 490  
475 TGATCGCTATGATGAGAGAGATGAGATGATGATGATGATGATGATGATGATGATGAT 534  
491 TAAATTAATGACATATCCAGTCCATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 550  
535 TAAATTAATGACATATCCAGTCCATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 594  
551 CCTTCTGGAAGCTCTTCAACCCAGAAATGGAATGAGAGAGTTCATTAATTGGCGCT 610  
595 CCTTCTGGAAGCTCTTCAACCCAGAAATGGAATGAGAGAGTTCATTAATTGGCGCT 654  
611 CAGACGAGTAACCTTGGCAAGGAGATGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 670  
655 CAGACGAGTAACCTTGGCAAGGAGATGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 714  
671 GGGGCTTTGTCGATCTCAAGCTGTCCGTAATCTCACAACCAAGATGCTTTGTGACAAC 730  
715 GGGGCTTTGTCGATCTCAAGCTGTCCGTAATCTCACAACCAAGATGCTTTGTGACAAC 774  
731 AAATGTCCTGTCAGATTTTCATCCCGTGAATCCGACCCGCGCTGACACTGTCTTAC 790  
775 AAATGTCCTGTCAGATTTTCATCCCGTGAATCCGACCCGCGCTGACACTGTCTTAC 834  
791 TCGGAACCTTGAAGAAACACCGGAGCTACCTGAGAGAGAGAGAGAGAGAGAGAGAG 850  
835 TCGGAACCTTGAAGAAACACCGGAGCTACCTGAGAGAGAGAGAGAGAGAGAGAGAG 894  
851 CATGAACCTTGAAGATCTCCGAGACACTAAACAGCACTTCACTGTGCGCTGACACT 910  
895 CATGAACCTTGAAGATCTCCGAGACACTAAACAGCACTTCACTGTGCGCTGACACT 954  
911 GGGATTCCTCACTCTT 928  
955 GGGATTCCTCACTCTT 972  
|||||

RESULT 7  
AL562459/c 972 bp mRNA linear EST 31-MAY-2003  
LOCUS AL562459 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens  
DEFINITION cDNA clone CS0DC005YN12 3-PRIME, mRNA sequence.

ACCESSION AL562459  
VERSION AL562459.2 GI:31286471  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
TITLE Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
JOURNAL Full-length cDNA libraries and normalization  
COMMENT On Feb 15, 2001 this sequence version replaced gi:12910899.  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 271.r For  
more information about this cluster, see  
http://www.genoscope.cns.fr/  
cgi-bin/cluster.cgi?seq=CS0DC005D606P1&cluster=271.r. Contact :  
Feng Liang Email : fliang@lifetech.com URL :  
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
Faraday Avenue Genoscope sequence ID : CS0DC005D606P1.  
Location/Qualifiers

FEATURES

```

source
1. .972
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DC005YN12"
/issue_type="NEUROBLASTOMA COT 25-NORMALIZED"
/clone_lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"
/notes="Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

```

## ORIGIN

```

Query Match      27.5%; Score 882.6; DB 9; Length 972;
Best Local Similarity 93.4%; Pred. No. 1.3e-218;
Matches 898; Conservative 37; Mismatches 24; Indels 2; Gaps 2;

QY 1251 AGGCTCAGAAAGAAAGAAATTGAGAGACAGAGGAAAGCCGCCACCGATGCTG 1310
    |||||
DB 972 AGGCTCAGAAAGAAAGAAATT-GAAGAAARARARARARARARARARARARAAAA 914
    |||||
QY 1311 CAGAAAGACAGAGGCGAGTCCGCTGCTGCGCTGATGAGGAATCTGTCGAGGAT 1370
    |||||
DB 913 MAAGAAAGAAAGAGGCGAGTCCAMMARACGCTGATGAGAAATCTGCCAAGGAT 854
    |||||
QY 1371 ATTGAGCAGAAAGATCCAGAAATGAGCTCATCATGCTCAATGCTGTAAGGAT 1430
    |||||
DB 853 ATTGAGCAGAAAGATCCAGAAATGAGCTCATCATGCTCAATGCTGTAAGGAT 794
    |||||
QY 1431 GTCAATGACAAAGCAGAGAAAGAGAGAGTGAAGTGAATGACGTAAGTCCCTG 1490
    |||||
DB 793 GTCAATGACAAAGCAGAGAAAGAGAGAGTGAAGTGAATGACGTAAGTCCCTG 734
    |||||
QY 1491 CAGTGCCTGAGTGAAGACTGGAAGCT-CATCTCAAGAGAGGCTCCAGAGGCTGGGCTGC 1549
    |||||
DB 733 CAGTGCCTGAGTGAAGACTGGAAGCTGGAAGCTGGAAGGCTCCAGAGGCTGGGCTGC 674
    |||||
QY 1550 TGGCTTTATGACCCGCTGTGAGGAGAAAGAAAGCTGAAGTCTATCAAGTCCG 1609
    |||||
DB 673 TAGATTTATTAACCCGCTGTGAGGAGAAAGAAAGCTGAAGTCTATCAAGTCCG 614
    |||||
QY 1610 GGGCTCTTCGATCAGGTGATGCTGTGACAGTGAAGGCTCCGATACCAAGAGATC 1669
    |||||
DB 613 GGGCTCTTCGATCAGGTGATGCTGTGACAGTGAAGGCTCCGATACCAAGAGATC 554
    |||||
QY 1670 CCAAGAGATGATACAGATGATTAATGCAAGAAACAGATTTTAAAGGCGGAGAA 1729
    |||||
DB 553 CCAAGAGATGATACAGATGATTAATGCAAGAAACAGATTTTAAAGGCGGAGAA 494
    |||||
QY 1730 AATCTTTCTGAGAGCTCAAAATTTGAAATGAAAAACCCAGACATCAGATGTTTT 1789
    |||||
DB 493 AATCTTTCTGAGAGCTCAAAATTTGAAATGAAAAACCCAGACATCAGATGTTTT 434
    |||||
QY 1790 ATTTTATATATATATATATAGAGGTGTACATTAATTAATGTAAGGAGATGACA 1849
    |||||
DB 433 ATTTTATATATATATATATAGAGGTGTACATTAATTAATGTAAGGAGATGACA 374
    |||||
QY 1850 CACCCAGCTTTGAGGGTCTGAGGGTAGAGCTGAGGAGCCCACTGGGAACAGACT 1909
    |||||
DB 373 CACCCAGCTTTGAGGGTCTGAGGGTAGAGCTGAGGAGCCCACTGGGAACAGACT 314
    |||||
QY 1910 GCAAGCTGAGCCATGAGCTGTTTTCCAAAGATCAGTTCTGAGGAGGAGGCTGAGCC 1969
    |||||
DB 313 GCAAGCTGAGCCATGAGCTGTTTTCCAAAGATCAGTTCTGAGGAGGAGGCTGAGCC 254
    |||||
QY 1970 TGACTCCGCTGTGTCCGAGACACAGTGTGACCGAGCCGCCCTGTATTTCTGG 2029
    |||||
DB 253 TGACTCCGCTGTGTCCGAGACACAGTGTGACCGAGCCGCCCTGTATTTCTGG 194
    |||||
QY 2030 CTGGGATCTGAGAGTGTGTGAGACACCTGAGCCCTACCAAGAGAGGTGAGGCACTTCT 2089
    |||||
DB 193 CTGGGATCTGAGAGTGTGTGAGACACCTGAGCCCTACCAAGAGAGGTGAGGCACTTCT 134
    |||||

```

```

QY 2090 GAGATCCAGCTGAACATGAGAAACAACTGAAAAAGCAGAGGCTCCGGTCAAGAG 2149
    |||||
DB 133 GAGATCCAGCTGAACATGAGAAACAACTGAAAAAGCAGAGGCTCCGGTCAAGAG 74
    |||||
QY 2150 CCTGTCTGTGTGCTGCTTCCATGACACACTCTCTCTGTGAATTAATCTGTAATC 2209
    |||||
DB 73 CCTGTCTGTGTGCTGCTTCCATGACACACTCTCTCTGTGAATTAATTAATTTTGAATC 14
    |||||
QY 2210 T 2210
    ||
DB 13 T 13

```

## RESULT 8

```

BO061784      1019 bp  mRNA  linear  EST 02-APR-2002
LOCUS         BO061784
DEFINITION    AGNCOURT_6854890 NIH_MGC_99 Homo sapiens cDNA clone IMAGE:5921279
5', mRNA sequence.
ACCESSION     BO061784
VERSION       BO061784.1 GI:19886171
KEYWORDS      EST.
SOURCE        Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 1019)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Lou Staudt
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LMC2086 row: e column: 24
High quality sequence strop: 596.

```

## FEATURES

```

source
1. .1019
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5921279"
/issue_type="lymphoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC_99"
/notes="Organ: Lymph; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGAG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
Library."

```

## ORIGIN

```

Query Match      27.3%; Score 876.4; DB 13; Length 1019;
Best Local Similarity 96.8%; Pred. No. 5.4e-217;
Matches 937; Conservative 0; Mismatches 26; Indels 5; Gaps 4;

QY 7 TTGCGAAGATGCGGACGCGCTTGACGAGAGAGACTGGAACAATGAAGCTATTACTGT 66
    |||||
DB 1 TTGCGAAGATGCGGACGCGCTTGACGAGAGAGACTGGAACAATGAAGCTATTACTGT 60
    |||||
QY 67 TGTGAAACGTGCGCAGGAGGCTCTTCTGAAAGAGCTGAAAGCTCTTACCGGAGCTCT 126
    |||||
DB 61 TGTGAAACGTGCGCAGGAGGCTCTTCTGAAAGAGCTGAAAGCTCTTACCGGAGCTCT 120
    |||||
QY 127 GTATGCTTACCATCCAGCAAGACAGAGACCAGAGCTCAAGTCAAGCGGAGAGAC 186
    |||||

```

[illegible]

**AUTHORS** NIH-MGC <http://mgc.nci.nih.gov/>.  
**TITLE** National Institutes of Health, Mammalian Gene Collection (MGC)  
**JOURNAL** Unpublished (1999)  
**COMMENT** Contact: Robert Strausberg, Ph.D.

Email: cgsabbs-r@mail.nih.gov  
 Tissue Procurement: Lou Staudt  
 CDNA Library Preparation: Rubin Laboratory  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: L1CM2040 row: C column: 09  
 High quality sequence scop: 629.

FEATURES	Location/Qualifiers
source	1. .1025

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5803712"
/tissue_type="lymphoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_95"
/notes="Organ: lymph; Vector: pOT87; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCGAG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
library."

```

**ORIGIN**

Query Match	27.0%;	Score 869;	DB 12;	Length 1025;
Best Local Similarity	98.3%;	Pred. No. 4.6e-215;		
Matches 941; Conservative	0;	Mismatches 10;	Indels 6;	Gaps 6;

Qy	7	TTGCGAAATGCGCAGCGGCTTTGAGCCGAGGAGAGGCTGGACAATGAAGAATAATTAATCTGT	66
Db	1	TTGCGAAGATGCGCAGCGGCTTTGAGCCGAGGAGAGCTGGACAATGAAGAATAATTAATCTGT	60
Qy	67	TGCTGAACGTGCGCAGGGAGGCTCTTCTGGAAGCTGAAAGCTGCTACCGAGGCTCT	126
Db	61	TGCTGAACGTGCGCAGGGAGGCTCTTCTGGAAGCTGAAAGCTGCTTACCGAGGCTCT	120
Qy	127	GTATGCTCTACATCATCCAGACAAGACACAGAGCCGACATCAATCCACAGCGGGAAGAC	186
Db	121	GTATGCTCTTACATCCAGACAAGACACAGAGCCGACATCAATCCACAGCGGGAAGAC	180
Qy	187	TGTTTTAACCTTGTTCACGAGGCTTATGAAAGTGTCTTATGTGACCCCCAAACGAGGCCATCT	246
Db	181	TGTTTTAACCTTGTTCACGAGGCTTATGAAAGTGTCTTATGTGACCCCCAAACGAGGCCATCT	240
Qy	247	ATGATATATATGAGGAAGAGAGACTGCGAAATGGAAGATGGAGAGTTGTGAAAGAGGA	306
Db	241	ATGATATATATGAGGAAGAGAGACTGCGAAATGGAAGATGGAGAGTTGTGAAAGAGGA	300
Qy	307	GAAACCCCTGCTGAATTTGAGAGAGAGTTTGACGCGCTTCAGAGAGAGAGAGAGAGGA	366
Db	301	GAAACCCCTGCTGAATTTGAGAGAGAGTTTGAGCGGCTTCAGAGAGAGAGAGAGAGGA	360
Qy	367	GATTGAGAGAGGAAACCAATCCCAAGGGAACGATCAGCGTTTGGAGTGTGATGCCACGACC	426
Db	361	GATTGAGAGAGGAAACCAATCCCAAGGGAACGATCAGCGTTTGGAGTGTGATGCCACGACC	420
Qy	427	TTTTTGGATGCGATATGATGAGAGATGAGAAAGTGTGTCGCGCAGTGAAGCTTTCCGCAATTG	486
Db	421	TTTTTGGATGCGATATGATGAGAGATGAGAAAGTGTGTCGCGCAGTGAAGCTTTCCGCAATTG	480
Qy	487	AAATTATATAAATGCAATATCCAGTCCATTTGAGGACCCCTTGACACGACACAGACAG	546
Db	481	AAATTATATATAAATGCAATATCCAGTCCATTTGAGGACCCCTTGACACGACACAGACAG	540



RESULT 11  
BUS38032 930 bp mRNA linear EST 13-SEP-2002  
LOCUS  
DEFINITION  
AGENCOURT 10186407 NIH\_MGC\_107 Homo sapiens cDNA clone  
IMAGE:6568542 5', mRNA sequence.  
ACCESSION  
BUS38032  
VERSION  
BUS38032.1 GI:22848473  
KEYWORDS  
EST.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE  
1 (bases 1 to 930)  
NIH-MGC <http://mgs.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
AUTHORS  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: ATCC  
COMMENT  
CDNA Library Preparation: Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.llnl.gov>  
Plate: LNCM2754 row: 9 column: 06  
High quality sequence stop: 673.  
Location/Qualifiers  
1..930  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6568542"  
/issue\_type="adenocarcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_1lb="NIH MGC 107"  
/note="Organ: breast; Vector: pORF7; Site 1: EcoRI;  
Site 2: XhoI; cDNA made by oligo-dT priming.  
Directionally cloned into EcoRI/XhoI sites using the  
following 5' adaptor: GGCACGAG(G). Library constructed by  
Ling Hong in the laboratory of Gerald M. Rubin (University  
of California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies).  
Note: this is a NIH\_MGC Library."

ORIGIN  
Query Match 26.8%; Score 861.6; DB 13; Length 930;  
Best Local Similarity 98.9%; Pred. No. 3.8e-213;  
Matches 908; Conservative 0; Mismatches 6; Indels 4; Gaps 4;  
QY 1353 GAATCTGTCCGAAGATTAATTGAGCGAAGAGTCCAGATGGCCTTCATCATGTCAAT 1412  
DB 1 GAAATCTGTCCGAAGAT-ATTGAGCGAGAGAGTCCAGAAATGGCCTTCATCATGTCAAT 59  
QY 1413 GCCGTGTACGGGAGTTTGTCAATGACACAGACGAGAGCGAAGAGTGAAGTATT 1472  
DB 60 GCTCTGTACGGGAGTTTGTCAATGACACAGAGCGAAGAGTGAAGTGAATT 119  
QY 1473 GACGTGACTGTGCCCTGTGACGTGCTGTGAAGACTGAGTCAATCTTCACGGAAGCC 1512  
DB 120 GACGTGACTGTGCCCTGTGACGTGCTGTGAAGACTGAGTCAATCTTCACGGAAGCC 179  
QY 1533 TCCAAAGCTGGGCTGCTGCTTTTATGACCCGCTGTGTGGGGGAAGAGAACTTGA 1592  
DB 180 TCCAAAGCTGGGCTGCTGCTTTTATGACCCGCTGTGTGGGGGAAGAGAACTTGA 239  
QY 1593 GTGCTCATCATGTTCCGGGGGCTCTGTCATCAGATGATGTTGCTGAGACAGTGAAGCCCTC 1652  
DB 240 GTGCTCATCATGTTCCGGGGGCTCTGTCATCAGATGATGTTGCTGAGACAGTGAAGCCCTC 299  
QY 1653 CGGATACCAAGACAGTCCACAGATGATGATGATGATGATGATGATGATGATGATGAT 1712  
DB 300 CGGATACCAAGACAGTCCACAGATGATGATGATGATGATGATGATGATGATGATGAT 359

QY 1713 TTTAAAGGCGCGAAAAATCTTTCTCGGAGCTTACAAATTTGGAATGAAAAAACC 1772  
DB 360 TTTAAAGGCGCGAAAAATCTTTCTCGGAGCTTACAAATTTGGAATGAAAAAACC 419  
QY 1773 AGACATCAGATGTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 1832  
DB 420 AGACATCAGATGTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 479  
QY 1833 GTGAAGGACATGACACACCCAGCTTTTGAAGGCTGCGGGGTGAGTGAAGTGAAGCC 1892  
DB 480 GTGAAGGACATGACACACCCAGCTTTTGAAGGCTGCGGGGTGAGTGAAGTGAAGCC 539  
QY 1893 CCACTGGGAACCAAGCTGACAGCTGCGGCCCATGCTGTTTCCCAAGATCATGTTCTGGA 1952  
DB 540 CCACTGGGAACCAAGCTGACAGCTGCGGCCCATGCTGTTTCCCAAGATCATGTTCTGGA 599  
QY 1953 GGGAAAGGCTCTGGGCTGACTCGCTGTGTCCGAGCACAGTGTGACCGGACGCCGC 2012  
DB 600 GGGAAAGGCTCTGGGCTGACTCGCTGTGTCCGAGCACAGTGTGACCGGACGCCGC 659  
QY 2013 CGCCTGTAGTCTTGGCTGAGTCTGTGAGAGTGTCTGTGAGCACCTTCCCTGACACAG 2072  
DB 660 CGCCTGTAGTCTTGGCTGAGTCTGTGAGAGTGTCTGTGAGCACCTTCCCTGACACAG 719  
QY 2073 GAGGTGAGGACCTTCTGACAGTCCAGCTGTAACATGGGAACAACCTGAAAAGCAGCAG 2132  
DB 720 GAGGTGAGGACCTTCTGACAGTCCAGCTGTAACATGGGAACAACCTGAAAAGCAGCAG 779  
QY 2133 GCCT-CCCGGTGAGGAGCCTGTGCTGTGCTGCTTCCATGACACCTCTCTGCTGA 2191  
DB 780 GCCTCCCCGTGAGGAGCCTGTGCTGTGCTGCTTCCATGACACCTCTCTGCTGA 839  
QY 2192 AATATTACTCTGTAATCTGAGACAGA-TTGCAGGTTTATTAACCTGTTTATCTGAG 2250  
DB 840 AA-ATTACTCTGTAATCTGAGACAGATTGGCGGNTTATTAACCTGTTTATCTGAG 898  
QY 2251 AACAAACGGGTTTGGAAA 2268  
DB 899 AACAAACGGGTTTGGAAA 916

RESULT 12  
BUI50592 935 bp mRNA linear EST 03-SEP-2002  
LOCUS  
DEFINITION  
AGENCOURT 8728400 NIH\_MGC\_47 Homo sapiens cDNA clone IMAGE:6339117  
5', mRNA sequence.  
ACCESSION  
BUI50592  
VERSION  
BUI50592.1 GI:22664124  
KEYWORDS  
EST.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE  
1 (bases 1 to 935)  
NIH-MGC <http://mgs.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
AUTHORS  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: ATCC  
COMMENT  
CDNA Library Preparation: Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.llnl.gov>  
Plate: LNCM2533 row: 9 column: 22  
High quality sequence stop: 675.  
Location/Qualifiers  
1..935  
/organism="Homo sapiens"  
/mol\_type="mRNA"







```

Db      118 CTGATGCTACCATCCAGCAAGACAGAGACCCAGAGCTCAAGTACAGGGGGAAAGC 177
Qy      185 ACTGTTTAACCTTGTTCACCGGCTTATGAAGTCTTATGTAACCCCAACACGAGGCCAT 244
Db      178 ACTGTTTAACCTTGTTCACCGGCTTATGAAGTCTTATGTAACCCCAACACGAGGCCAT 237
Qy      245 CTATGATATATATGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 304
Db      238 CTATGATATATATGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 297
Qy      305 GAGAACCCCTGCTGAAATTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 364
Db      298 GAGAACCCCTGCTGAAATTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 357
Qy      365 GAGATTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 424
Db      358 GAGATTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 417
Qy      425 CCTTTTGTATGCTATGATGAGAGATGATGAGAGATGATGAGAGATGATGAGAGATGATGAGAG 484
Db      418 CTTTGTGATGCTATGATGAGAGATGATGAGAGATGATGAGAGATGATGAGAGATGATGAGAGAT 477
Qy      485 TGAATTAATAAATGACATATCCAGTCCATTCAGGCAACCTTGAACAGAGACAGACAC 544
Db      478 TGAATTAATAAATGACATATCCAGTCCATTCAGGCAACCTTGAACAGAGACAGACAC 537
Qy      545 AGCCATCTCTCTGGAAGGCTTCAACCCAGAGATGGAATGAGAGAGAGAGAGAGAGAGAGAG 604
Db      538 AGCCATCTCTCTGGAAGGCTTCAACCCAGAGATGGAATGAGAGAGAGAGAGAGAGAGAGAG 597
Qy      605 TGGGCTGAGAGAGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 664
Db      598 TGGGCTGAGAGAGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 657
Qy      665 CCTACAGAGGAGCTTGTGCTGCTCAAGGCTTCCGATATTCACACCAAGATGCTTGT 724
Db      658 CCTACAGAGGAGCTTGTGCTGCTCAAGGCTTCCGATATTCACACCAAGATGCTTGT 717
Qy      725 GACAAACAACTGTGCTGCTCAAGGCTTCCGATATTCACACCAAGATGCTTGT 784
Db      718 GACAAACAACTGTGCTGCTCAAGGCTTCCGATATTCACACCAAGATGCTTGT 777
Qy      785 CCTAGCTCGAGACCTAGACAG- AACCACCGGAGCTACCTGACAGTGGGAT- GGGGATAC 842
Db      778 CTTAGCTCGAGACCTAGACAG- AACCACCGGAGCTACCTGACAGTGGGAT- GGGGATAC 837
Qy      843 CAGTCAAGCAGT- AACACTAGATC- GTCCGAGACACTAAACAGGCACTTCACTGTGG 900
Db      838 CCGTCAAGCAGTAAACACTAGATC- GTCCGAGACACTAAACAGGCACTTCACTGTGG 897
Qy      901 - -CCCTGAGCTGGAATTCCTCACTCTCTTGTGACT- GATCAGCTATACAGCAAAATTC 957
Db      898 GCCCTGAGCTGGAATTCCTCACTCTCTTGTGACTGATCAGCTATACAGCAAAATTC 957
Qy      958 AAGATGAC- -GATCAGCTGCTGGAAGATCCTTCAAGGAGCTTCTTGGG 1010
Db      958 AAGATGACCGATCCGATCCGCTGGAAGATCCTTCAAGGAGCTTCTTGGG 1013

```

```

RESULT 14
LOCUS    B0063825 1018 bp mRNA linear EST 02-APR-2002
DEFINITION AGENTCOURT_6850301 NIH_MGC_99 Homo sapiens cDNA clone IMAGE:5925479
5', mRNA sequence.
ACCESSION B0063825
VERSION   B0063825.1 GI:19891913
KEYWORDS  EST.
SOURCE    Homo sapiens (human)
ORGANISM  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1018)

```

```

AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE    National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL  Unpublished (1999)
COMMENT  Contact: Robert Strausberg, Ph.D.
         Email: sgabbs-remail.nih.gov
         Tissue Procurement: Lou Staudt
         cDNA Library Preparation: Rubin Laboratory
         cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
         DNA Sequencing by: Agencourt Bioscience Corporation
         Clone distribution: MGC clone distribution information can be
         found through the I.M.A.G.E. Consortium/LNL at:
         http://image.lnl.gov
         plate: LNCM2097 row: d column: 24
         High quality sequence stop: 758.
FEATURES
Source
Location/Qualifiers
1..1018
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5925479"
/tissue_type="lymphoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_99"
/notes="Organ: lymph; Vector: pOTB7; Site 1: XhoI; Site 2:
into EcoRI/XhoI sites using the following 5' adaptor:
GGCACGAG(G). Size selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
Library."
ORIGIN
Query Match 26.7%; Score 856.6; DB 13; Length 1018;
Best Local Similarity 96.8%; Pred. No. 7.9e-212;
Matches 928; Conservative 0; Mismatches 21; Indels 10; Gaps 5;
Qy      7 TTGGGAAGATGGGAGCGGCTTGAAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 66
Db      1 TTGGGAAGATGGGAGCGGCTTGAAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 60
Qy      67 TGGTGAACGTCGAGAGAGAGCGCTTCTGAAGAGCTGAAGAGCTTACCGGAGGCT 126
Db      61 TGGTGAACGTCGAGAGAGAGCGCTTCTGAAGAGCTGAAGAGCTTACCGGAGGCT 120
Qy      127 GTATGCTCTACATCCAGACAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 186
Db      121 GTATGCTCTACATCCAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
Qy      187 TGTTTAACTTGTTCACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 246
Db      181 TGTTTAACTTGTTCACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240
Qy      247 ATGATATATATGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 306
Db      241 ATGATATATATGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
Qy      307 GAAACCCCTGCTGAAATTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 366
Db      301 GAAACCCCTGCTGAAATTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
Qy      367 GATTGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 426
Db      361 GATTGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
Qy      427 TTTTGTATGCTATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 486
Db      421 TTTTGTATGCTATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
Qy      487 AAATTAATAAATGACATATCCAGTCCATTCAGGCAACCTTGAACAGAGAGAGAGAGAGAGAG 546
Db      481 AAATTAATAAATGACATATCCAGTCCATTCAGGCAACCTTGAACAGAGAGAGAGAGAGAGAG 540

```

QY 547 CCATCTCTCTGGAAGCCTCTCAACCCAGATGAAATGAGGAGTTCCATTACCTTG 606  
 541 CCATCTCTCTGGAAGCCTCTCAACCCAGATGAAATGAGGAGTTCCATTACCTTG 600  
 QY 607 CGCTCAGACGAGTAACTTCCGCAAGAGATGAGGAGTTGAAATTTGAGCTGAGACC 666  
 601 CGCTCAGACGAGTAACTTCCGCAAGAGATGAGGAGTTGAAATTTGAGCTGAGACC 660  
 QY 667 TAAGAGGAGCCTTTGTTCCGCTCAAGCTGTTCCGTAATCTCAACCAAGATGCTTTGTGA 726  
 661 TAAGAGGAGCCTTTGTTCCGCTCAAGCTGTTCCGTAATCTCAACCAAGATGCTTTGTGA 720  
 QY 727 CAACAACTGTGCTGTCAGATTTTCAATCCGCTGGAATCCGACCCGCTGACCACTGTCC 786  
 721 CAACAACTGTGCTGTCAGATTTTCAATCCGCTGGAATCCGACCCGCTGACCACTGTCC 780  
 Db 787 TAGCTCGGAACCTAGACAAAGAACACCGT-GAGGCTACCTGCAATGAGCGAT-GAGGTAATCCA 844  
 781 TAGCTCGGAACCTAGACAAAGAACACCGTGGGCTACCTGCAATGAGCGATGAGGTAATCCA 840  
 QY 845 GTAGCGCAAGAACATAGCATGCTCCGAGACACTAAACCC-AGCCACTTCACTGTGGCC 902  
 841 GTAGCGCAAGAACATAGCATGCTCCGAGACACTAAACCCAGCCACTTCACTGTGGCC 900  
 Db 903 C-TGCAGCTGGAA-----TCCCTCACTCTTTGCACTGATCAGCTATCAGCAAAATT 955  
 901 CGTGACGCTGGAAATCCCTCACTCTCTTTTGCACTGATCAGCTATCAGCAAAAT 959

RESULT 15  
 AL575646/c 989 bp mRNA linear EST 01-JUN-2003  
 LOCUS AL575646 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA  
 DEFINITION clone CS0D1069Y21 3-PRIME, mRNA sequence.  
 ACCESSION AL575646  
 VERSION AL575646.2 GI:31313953  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 989)  
 Li, W.B., Gruber, C., Jesse, J., and Polayes, D.  
 Full-length cDNA libraries and normalization  
 Unpublished (2001)  
 On Feb 16, 2001 this sequence version replaced gi:12937012.  
 CONTACT: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 EVRY cedex - France  
 Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr  
 Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 271.r For more information about this cluster, see  
 http://www.genoscope.cns.fr/  
 cgi-bin/cluster.cgi?seq=CS0D1069CH1NP1&cluster=271.r. Contact :  
 Feng Liang Email: fliang@lifetech.com URL :  
 http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
 Faraday Avenue Genoscope sequence ID : CS0D1069CH1NP1.

FEATURES  
 source  
 1. 989  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0D1069Y21"  
 /tissue\_type="PLACENTA COT 25-NORMALIZED"  
 /clone\_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"  
 /note="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end cloned, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match 26.5%; Score 852.8; DB 9; Length 989;  
 Best local Similarity 95.2%; Pred. No. 7.7e-211;  
 Matches 902; Conservative 13; Mismatches 28; Indels 4; Gaps 4;

QY 2238 CTTTATCTGAGAACAAACGGGTTTGGAAATTAAGTCTTTTCCCACTCCAGAG 2297  
 945 CTGTTTATCTGAGAACAAACGGGTTTGGAAATTAAGTCTTTTCCCACTCCAGAG 887  
 Db 2298 CTGCTCAAGTCAATTCACACCGGCCCCCTGGGCTTTGGGACAGGGTATGTAATCTCCGATCC 2357  
 886 CTGCTCAAGTCAATTCACACCGGCCCCCTGGGCTTTGGGACAGGGTATGTAATCTCCGATCC 827  
 QY 2358 CAGGCGCTAGCCCTGACACAGGTGCTTCCGTAATCCCGTGGGAAAAGCCCTCGCAC 2417  
 826 CA-GGCTAGCCCTGACACAGGTGCTTCCGTAATCCCGTGGGAAAAGCCCTCGCAC 768  
 Db 2418 AGCGGCTTAGCTGAGCTGTGTCTCTCCACCGCTGACACCACTCCAGAGTGCAG 2477  
 767 AGCGGCTTAGCTGAGCTGTGTCTCTCCACCGCTGACACCACTCCAGAGTGCAG 708  
 QY 2478 TGCTGGGCAAGGGCAGCTCAAGAGACAGGACGAGGCGCTTGGCAAGATGACACAC 2537  
 707 TGCTGGGCAAGGGCAGCTCAAGAGACAGGACGAGGCGCTTGGCAAGATGACACAC 648  
 Db 2538 CCAACCCAAAGGCGTGAACCCCAAGGCCCCGCTGTGTAACCAAGAGGTGACATGACAGC 2597  
 647 CCAACCCAAAGGCGTGAACCCCAAGGCCCCGCTGTGTAACCAAGAGGTGACATGACAGC 588  
 QY 2598 TCCCGGCTCTGCAAGGTCCAGGCTCTCAAGAAACCAAGGCGCTGTGCTCCGAGCCT 2657  
 587 TCCCGGCTCTGCAAGGTCCAGGCTCTCAAGAAACCAAGGCGCTGTGCTCCGAGCCT 528  
 Db 2658 TCCCTCAGACCCCTCCTCCACAGTGCACCTTGGAGTGCAGAAATGACAGGAGCTAGAGC 2717  
 527 TCCCTCAGACCCCTCCTCCACAGTGCACCTTGGAGTGCAGAAATGACAGGAGCTAGAGC 468  
 QY 2718 CCTCCACGCGCTGGAACCTCGGCTGCAATTAAGTAACTGAGGCTGTCTCGGGGCT 2777  
 467 CCTCCACGCGCTGGAACCTCGGCTGCAATTAAGTAACTGAGGCTGTCTCGGGGCT 408  
 Db 2778 GGAATGGGACGCAATCAGTGTCTCTTGTGACCCCTCGAGGACAGCGCCGACAGGTGT 2837  
 407 GGAATGGGACGCAATCAGTGTCTCTTGTGACCCCTCGAGGACAGCGCCGACAGGTGT 348  
 QY 2838 GGCTGAGACAGTGGGCGGGGGGCCCCAAGGTGGCGGCTCCAGGCCACCAAGCT 2897  
 347 GGCTGAGACAGTGGGCGGGGGGCCCCAAGGTGGCGGCTCCAGGCCACCAAGCT 289  
 Db 2898 GTTGTGAAGTCAAGGCAACTCCCGACGACTGTATCTGATTAACGGCTAAGAACTCC 2957  
 288 GTTGTGAAGTCAAGGCTCCCTCCCGACGACTGTATCTGATTAACGGCTAAGAACTCC 229  
 QY 2958 TTCTCTGTGTTTGAAGACAGTGGGTTGTCTCAATCTGTAACTTATCTCAATTT 3017  
 228 TTCTCTGTGTTTGAAGACAGTGGGTTGTCTCAATCTGTAACTTATCTCAATTT 169  
 Db 3018 TTAATAAAGGTTTCTGACGCGCCCAAGGCGGCGGCTGAGCGGTGTTGACATG 3077  
 168 TT-TAATAAAGTCTGACGCGCCCAAGGCGGCGGCTGAGCGGTGTTGACATG 110  
 QY 3078 AGCTGAGGCGCGGAGCTTCCCGTGGCGCTTCCCGACAGGTGTTTGGGACCAATCTC 3137  
 109 AGCTGAGGCGCGGAGCTTCCCGTGGCGCTTCCCGACAGGTGTTTGGGACCAATCTC 50  
 Db 3138 TCGCTTCAATTTGCACTGACGTGTACAGAGGCACTCACCAATTA 3184  
 49 TCGCTTCAATTTGCACTGACGTGTACAGAGGCACTCACCAATTA 3

Search completed: July 13, 2004, 14:50:21  
 Job time : 768 secs